مقدمة فى
أساليب الإستدلال الإحصائى والتنبؤ

الأستاذة الدكتورة
إمتثال محمد حسن
كلية التجارة - جامعة الأسكندرية

الأستاذ الدكتور
عادل محمود حلاوة
كلية التجارة - جامعة الأسكندرية

الدكتورة
لبيبة حسب النبى العطار
كلية التجارة - جامعة الأسكندرية



٠٠٢٠١٠٠٣٧٣٨٨٢٢

مقدمة فى

أساليب الإستدلال الإحصائى والتنبؤ

الأستاذة الدكتورة
إمتثال محمد حسن
كلية التجارة جامعة الإسكندرية

الأستاذ الدكتور
عادل محمود حلاوة
كلية التجارة جامعة الإسكندرية

الدكتورة
لبيبة حسب النبى العطار
كلية التجارة جامعة الإسكندرية

الطبعة الأولى
2012م

الناشر
مكتبة الوفاء القانونية
محمول : 0020103738822 الإسكندرية

مقدمة

إن التطور التكنولوجي الحديث في جميع مجالات حياتنا المعاصرة من ناحية ، ودخول العالم في عصر المعلوماتية ، من ناحية أخرى ، كل هذا أدى إلى ازدياد أهمية استخدام أساليب التحليل الإحصائي في جميع مجالات المعرفة ، وعلى جميع المستويات . فعلى مستوى الاقتصاد القومي ، أو مستوى الوحدات الاقتصادية ، سواء كانت قطاع عام أو خاص ، فإن الاحتياج إلى جمع البيانات واستخراج المعلومات بناءً على أساس من الدراسة المنهجية الحديثة ، يعتبر من المسائل الحيوية في عصرنا الحديث . وهو ما تقوم به أساليب التحليل الإحصائي . وتستخدم هذه الأساليب في التخطيط ورسم السياسات واتخاذ القرارات والتنبؤات .

ولقد تناول هذا الكتاب " مقدمة في أساليب التحليل الإحصائي " بعرض أهم هذه الأساليب وكيفية الوصول منها إلى نتائج وقرارات إحصائية تفيد الباحث ومتخذ القرار . وتتطلب دراسة هذه الموضوعات الإلمام بمبادئ الإحصاء الوصفي . ولقد روعي التبسيط في عرض المواضيع دون الإخلال بالمادة العلمية . ألا أن هذا الكتاب لم يتعرض لاستخدام الحاسبات الآلية في التحليل الإحصائي ، والسبب في ذلك يرجع إلى أن الكتاب مقرر على السنة الثانية بكلية التجارة ، قبل التخصص ، ونظراً للأعداد الكبيرة من الطلبة فإن الإمكانيات لا تسمح باستخدام التطبيقات على الحاسب الآلي . ألا أن هناك مقرراً في التطبيقات الإحصائية على الحاسب الآلي لطلبة التخصص في الإحصاء التطبيقي في السنة الثالثة ، حيث أعداد الطلبة صغيرة .

هذا وقد قام بتأليف هذا الكتاب كل من أ.د. امتثال محمد حسن ، أ.د. عادل محمود حلاوة و د. لبيبة حسب النبي العطار.

قامت أ.د. امتثال محمد حسن بكتابة الفصول الثلاثة الأولى.

قام أ.د. عادل محمود حلاوة بكتابة الفصل الرابع والفصلين السابع والثامن.

قامت د. لبيبة حسب النبي العطار بكتابة الفصلين الخامس والسادس.

هذا ويتمنى مؤلفو الكتاب لأبنائهم الطلبة حسن الأداء خلال الفصل الراسي حتى تكلل مجهوداتهم بالنجاح.

الإسكندرية في يناير ٢٠١١

المؤلفون

الفصل الأول

توزيعات المعاينة

مقدمة :

قد يكون من المفيد ، بادئ ذي بدء ، تذكرة الطالب ببعض التعريفات الهامة التي سيحتاجها في دراسته لهذا الفصل . ومن هذه التعريفات : تعريف المجتمع Population وتعريف العينة Sample .

ويمكن تعريف المجتمع بأنه " جميع " المفردات محل الدراسة سواء كانت في شكل إنسان أو حيوان أو جماد أو أشياء غير ملموسة ، وسواء كان من الممكن عدها أم لا . فيقال مثلاً : مجتمع درجات الطلبة في امتحان الإحصاء ، وهذا يعني درجات جميع الطلبة الذين دخلوا امتحان الإحصاء . أما العينة فيمكن تعريفها بأنها " مجموعة " من مفردات المجتمع . ففي مثالنا السابق إذا كان عدد الطلبة الذين تقدموا لامتحان الإحصاء ١٠٠٠ طالب ، فالمجتمع هنا هو درجات هؤلاء الطلبة الألف في مادة الإحصاء . وإذا أخذت درجات ٥٠ طالب فقط من الذين تقدموا لهذا الامتحان نقول أنه تم اختيار عينة من درجات ٥٠ طالب من مجتمع درجات الإحصاء .

وقد تكون العينة عينة عشوائية بسيطة Simple random sample حيث تكون لكل مفردة من مفردات المجتمع نفس الفرصة في تكوين العينة .

وفي أغلب الأحيان يطلق على هذا النوع " العينة العشوائية " random sample . وقد تكون العينة غير عشوائية nonrandom sample حيث لا يكون لبعض المفردات نفس الفرصة في تكوين العينة . ولإيضاح ذلك فإذا كان لدينا مجتمع مكون من ١٠٠ شخص ولدينا خمس جوائز فقط . فإذا وضعنا أسماء الأشخاص المائة في سلة وسحبنا منها خمس أسماء ، نكون إزاء عينة عشوائية . أما إذا كتبنا أسماء المائة شخص حسب الترتيب الأبجدي واخترنا الخمس أسماء الأولى ، فإن العينة تكون غير عشوائية لأن باقي الأشخاص ـ وهم ٩٠ شخص ـ لا تكون لهم فرصة في تكوين العينة .

وتسمى دراسة جميع مفردات المجتمع " بالحصر الشامل " census وهو المستخدم في التعدادات السكانية والزراعية والصناعية ... الخ .إلا أن هذا الأسلوب يتطلب تكاليف باهظة ويستهلك كثيراً من الوقت والجهد ، وقد يؤدي هذا الأسلوب إلى تلف الوحدات محل الدراسة . لذلك يلجأ كثير من الباحثين إلى استخدام أسلوب " المعاينة الإحصائية " أي أسلوب استخدام العينات .

هذا ويسمى أي مقياس إحصائي في المجتمع " بالمعلمة " parameter في حين أن أي مقياس إحصائي في العينة يسمى " إحصائية " statistic . وعادة نرمز لمعالم المجتمع بالحروف الإغريقية مثال ذلك الوسط الحسابي في المجتمع يرمز له بالرمز μ ، والانحراف المعياري في المجتمع بالرمز σ ، والنسبة في المجتمع بالرمز θ . ويجدر الإشارة هنا أن معالم المجتمع دائماً ثابتة في حين أن إحصائيات العينة فهي دائماً متغير عشوائي random variable .

وقد تكون إحصائية العينة مساوية لمعلمة المجتمع ، أو قد تكون أصغر أو أكبر منها . ويسمى الفرق بين إحصائية العينة ومعلمة المجتمع بخطأ المعاينة sampling error أو الخطأ العشوائي random error ، وهذا بافتراض أن العينة عشوائية وأنه لا يوجد أخطاء غير عشوائية . ويقصد بالأخطاء غير العشوائية الأخطاء في تجميع البيانات وتدوينها وتبويبها .

وبما أن الفرق بين إحصائية العينة ومعلمة المجتمع قد يكون سالباً أو موجباً ، لذلك يمكننا أخذ القيمة المطلقة لهذا الفرق لحساب خطأ المعاينة . أي أن :

خطأ المعاينة = | إحصائية العينة – معلمة المجتمع | (١)

ويقوم هذا الفصل بدراسة توزيعات المعاينة : فينقسم إلى خمس مباحث رئيسية . يتناول المبحث الأول توزيع المجتمع ، ويتناول المبحث الثاني توزيع معاينة الوسط الحسابي $\bar{س}$ في حالة السحب بإرجاع والسحب بدون إرجاع ، ويتناول المبحث الثالث الوسط الحسابي والانحراف المعياري لتوزيع معاينة $\bar{س}$ ، ويتناول المبحث الرابع شكل توزيع معاينة $\bar{س}$ ، ويتناول المبحث الخامس توزيع معاينة النسبة ق .

(١ ـ ١) توزيع المجتمع :

بالتعريف توزيع المجتمع هو التوزيع الاحتمالي لجميع مفردات المجتمع . ويبين المثال التالي كيفية الحصول على توزيع المجتمع .

مثال (١) :

في أحد مكاتب المحاسبة يوجد ٥ موظفين ، وفيما يلي عدد سنوات الخبرة لهؤلاء الموظفين :

٦ ، ٤ ، ١٥ ، ٦ ، ١٠

والمطلوب : إيجاد التوزيع الاحتمالي لسنوات الخبرة في هـــذا المجتمـــع ومنها حساب الوسط الحسابي والانحراف المعيـــاري لـــهذا المجتمع .

الحـــل :

من الملاحظ أن القيمة ٦ تكررت مرتين . ويبين جـــدول (١) التوزيع التكراري لسنوات الخبرة .

جدول (١)

التوزيع التكراري لسنوات الخبرة

سنوات الخبرة س	التكرار ك .	س ك	$س^2$ ك
٤	١	٤	١٦
٦	٢	١٢	٧٢
١٠	١	١٠	١٠٠
١٥	١	١٥	٢٢٥
المجموع	**٥**	**٤١**	**٤١٣**

ومن هذا التوزيع يمكن حساب الوسط الحســابي والانحــراف المعيــاري كما يلي :

الوسط الحسابي للمجتمع : $\mu = \frac{\text{مجـ س ك}}{\text{مجـ ك}} = \frac{٤١}{٥} = ٨,٢$ سنة

تباين المجتمع : $\sigma^{٢} = \frac{١}{\text{مجـ ك}} [\text{مجـ س}^{٢} \text{ك} - \frac{(\text{مجـ س ك})^{٢}}{\text{مجـ ك}}]$

$= \frac{١}{٥} [٤١٣ - \frac{(٤١)^{٢}}{٥}] = ١٥,٣٦$

∴ الانحراف المعياري $\sigma = \sqrt{١٥,٣٦} = ٣,٩١٩$ سنة

ومن التوزيع التكراري يمكننا الحصول على التوزيع الاحتمالي عن طريق إيجاد التكرار النسبي بقسمة التكرار على مجموع التكرارات . ويبين الجدول التالي التوزيع الاحتمالي لسنوات الخبرة :

جدول (٢)

التوزيع الاحتمالي لسنوات الخبرة

سنوات الخبرة س	التكرار النسبي ح (س)	س . ح (س)	س^{٢} . ح (س)
٤	$\frac{١}{٥}$ = ٠,٢٠	٠,٨	٣,٢
٦	$\frac{٢}{٥}$ = ٠,٤٠	٢,٤	١٤,٤
١٠	$\frac{١}{٥}$ = ٠,٢٠	٢	٢٠
١٥	$\frac{١}{٥}$ = ٠,٢٠	٣	٤٥
المجموع	١٠٠	٨,٢	٨٢,٦

ومن هذا التوزيع يمكن الحصول على الوسط الحسابي والانحراف المعياري كما يلي :

الوسط الحسابي للمجتمع : μ = مجـ س . ح (س) = ٨,٢ سنة

تباين المجتمع : σ^2 = مجـ $س^2$. ح (س) − μ^2

= ٨٢,٦ − $(٨,٢)^2$ = ١٥,٣٦

∴ الانحراف المعياري σ = $\sqrt{١٥,٣٦}$ = ٣,٩١٩ سنة

ومن الواضح أن الوسط الحسابي والانحراف المعياري لهما نفس القيمة سواء كانا حسبا من التوزيع التكراري أو التوزيع الاحتمالي .

(١ ــ ٢) توزيع المعاينة للوسط الحسابي $\bar{س}$ في حالة السحب بإرجاع وبدون إرجاع :

رأينا في المبحث السابق كيفية الحصول على الوسط الحسابي للمجتمع μ . وكما سبق وذكرنا ، فإن الوسط الحسابي للمجتمع هو مقدار دائما ثابت ، أما الوسط الحسابي في العينة فتختلف قيمته من عينة إلى أخرى مسحوبة من نفس المجتمع ؛ ومن ثم فإن الوسط الحسابي متغيرا عشوائيا . وبما أن الوسط الحسابي في العينة متغيرا عشوائيا ، فإن له توزيع احتمالي يسمى بتوزيع المعاينة أو التوزيع العيني sampling distribution للوسط الحسابي $\bar{س}$. ونحصل على توزيع المعاينة بأخذ جميع العينات الممكنة التي لها نفس الحجم ، والتي يمكن الحصول عليها من هذا المجتمع . وفيما يلي سنتعرض لتوزيع معاينة $\bar{س}$ في حالة السحب بإرجاع وفي حالة السحب بدون إرجاع ، من خلال المثال التالي :

مثال (٢) :

بأخذ المجتمع الموجود في مثال (١) والخاص بعدد سنوات الخبرة لموظفي أحد مكاتب المحاسبة وعددهم ٥ ، وبوضع رمز لكل مفردة من هذه المفردات سيكون لدينا ما يلي :

سنوات الخبرة	:	٦	٤	١٥	٦	١٠
الرمز	:	أ	ب	حـ	د	هـ

والمطلوب :

١ ــ إيجاد جميع العينات الممكنة من مفردتين التي يمكن سحبها من هذا المجتمع بإرجاع ، وحساب خطأ المعاينة ، ثم إيجاد توزيع المعاينة في هذه الحالة ومنه حساب الوسط الحسابي والانحراف المعياري .

٢ ــ إيجاد جميع العينات الممكنة المكونة من مفردتين التي يمكن سحبها من هذا المجتمع بدون إرجاع ، ثم إيجاد توزيع المعاينة في هذه الحالة ومنه حساب الوسط الحسابي والانحراف المعياري للتوزيع .

١ ــ حالة السحب بإرجاع :

في حالة السحب بإرجاع فإن حجم المجتمع م لا يتغير عند سحب مفردة من مفردات المجتمع ، وهذا يعني أن سحب أي مفردة لا يتأثر بسحب المفردات السابقة ، ومن ثم تكون المشاهدات مستقلة عن بعضها البعض . ويمكن سحب المفردة أكثر من مرة لتكوين العينة .

فإذا أخذنا جميع العينات الممكنة المكونة من مفردتين التي يمكن سحبها من هذا المجتمع بإرجاع ، فإن :

عدد العينات = $م^{ن}$

حيث : م حجم المجتمع ، ن حجم العينة .

∴ $م^ن$ = $٥^٢$ = ٢٥ عينة .

ويبين جدول (٣) هذه العينات والوسط الحسابي لكل منها .

جدول (٣)

جميع العينات الممكنة في حالة السحب بإرجاع

والوسط الحسابي لكل منها

العينة	مفردات العينة	الوسط الحسابي $\bar{س}$
أ ، أ	٦ ، ٦	٦
أ ، ب	٦ ، ٤	٥
أ ، حـ	٦ ، ١٥	١٠,٥
أ ، د	٦ ، ٦	٦
أ ، هـ	٦ ، ١٠	٨
ب ، أ	٤ ، ٦	٥
ب ، ب	٤ ، ٤	٤
ب ، حـ	٤ ، ١٥	٩,٥
ب ، د	٤ ، ٦	٥
ب ، هـ	٤ ، ١٠	٧
حـ ، أ	١٥ ، ٦	١٠,٥
حـ ، ب	١٥ ، ٤	٩,٥
حـ ، حـ	١٥ ، ١٥	١٥
حـ ، د	١٥ ، ٦	١٠,٥
حـ ، هـ	١٥ ، ١٠	١٢,٥
د ، أ	٦ ، ٦	٦
د ، ب	٦ ، ٤	٥
د ، حـ	٦ ، ١٥	١٠,٥
د ، د	٦ ، ٦	٦
د ، هـ	٦ ، ١٠	٨
هـ ، أ	١٠ ، ٦	٨
هـ ، ب	١٠ ، ٤	٧
هـ ، حـ	١٠ ، ١٥	١٢,٥
هـ ، د	١٠ ، ٦	٨
هـ ، هـ	١٠ ، ١٠	١٠
المجموع		**٢٠٥**

ومن هذا الجدول حصلنا على الوسط الحسابي بقسمة مفردات العينة على عددها ، فمثلاً عندما كانت المفردات (٦ ، ٦) فإن :

$$\bar{س} = \frac{٦ + ٦}{٢} = ٦$$

ومن هذا الجدول يتضح لنا أن الأوساط الحسابية تختلف قيمتها من عينة إلى أخرى وفقاً لمفردات العينة ، فهو إذن متغيراً عشوائياً كما سبق وذكرنا . ومن هذا الجدول يتضح لنا أيضاً أن هذه الأوساط الحسابية تختلف عن الوسط الحسابي للمجتمع μ . ومن ثم يمكن حساب خطأ المعاينة للوسط الحسابي كما يلي :

$$\text{خطأ المعاينة للوسط الحسابي} = |\bar{س} - \mu| \qquad (٢)$$

فمثلاً بالنسبة للعينة الأولى (أ ، أ) :

خطأ المعاينة = | ٦ ـ ٨,٢ | = ٢,٢ سنة

وبالنسبة للعينة الثانية (أ ، ب) :

خطأ المعاينة = | ٥ ـ ٨,٢ | = ٣,٢ سنة

وهكذا بالنسبة لباقي العينات .

وإذا أخذنا الوسط الحسابي لهذه المتوسطات نحصل على ما يلي :

$$\bar{\bar{س}} = \frac{٢٠٥}{٢٥} = ٨,٢ \text{ سنة}$$

ومن جدول (٣) يمكننا الحصول على التوزيع التكراري للأوساط الحسابية $\bar{س}$ عندما تكون العينة مكونة من مفردتين في حالة السحب بإرجاع ، كما هو موضح في جدول (٤) .

جدول (٤)

التوزيع التكراري للأوساط الحسابية (س̄) للعينات

عندما تكون العينة مكونة من مفردتين في حالة السحب بإرجاع

الوسط الحسابي س̄	٤	٥	٦	٧	٨	٩,٥	١٠	١٠,٥	١٢,٥	١٥	المجموع
التكرار ك	١	٤	٤	٢	٤	٢٠	١	٤	٢	١	٢٥

وبقسمة التكرار على مجموع التكرارات نحصل على التكرار النسبي ، ثم نحصل على التوزيع الاحتمالي وهو في هذه الحالة توزيع المعاينة للأوساط الحسابية . كما هو مبين في جدول (٥) .

جدول (٥)

توزيع المعاينة للأوساط الحسابية (س̄)

عندما تكون العينة مكونة من مفردتين في حالة السحب مع الإرجاع

الوسط الحسابي س̄	٤	٥	٦	٧	٨	٩,٥	١٠	١٠,٥	١٢,٥	١٥	المجموع
التكرار النسبي ح (س̄)	١/٢٥	٤/٢٥	٤/٢٥	٢/٢٥	٤/٢٥	٢/٢٥	١/٢٥	٤/٢٥	٢/٢٥	١/٢٥	١

ولقد سبق وبينا في مثال (١) أنه يمكن حساب الوسط الحسابي والتباين من التوزيع التكراري أو التوزيع الاحتمالي . وفيما يلي سنستخدم التوزيع التكراري لحساب كل من الوسط الحسابي والتباين .

جدول (٦)

حساب الوسط الحسابي والانحراف المعياري

لتوزيع معاينة $\bar{س}$ في حالة السحب بإرجاع

الوسط الحسابي $\bar{س}$	التكرار ك	$\bar{س}$ ك	$\bar{س}^{٢}$ ك
٤	١	٤	١٦
٥	٤	٢٠	١٠٠
٦	٤	٢٤	١٤٤
٧	٢	١٤	٩٨
٨	٤	٣٢	٢٥٦
٩,٥	٢	١٩	١٨٠,٥
١٠	١	١٠	١٠٠
١٠,٥	٤	٤٢	٤٤١
١٢,٥	٢	٢٥	٣١٢,٥
١٥	١	١٥	٢٢٥
المجموع	٢٥	٢٠٥	١٨٧٣

ومن جدول (٦) يمكننا الحصول على الوسط الحسابي لتوزيع المعاينة ، والذي يرمز له بالرمز ($\mu_{\bar{س}}$) ، كما يلي :

$$\mu_{\bar{س}} = \frac{\text{مجـ } \bar{س} \text{ ك}}{\text{مجـ ك}} = \frac{٢٠٥}{٢٥} = ٨,٢ \text{ سنة}$$

ولقد حصلنا على نفس هذه القيمة بحساب الوسط الحسابي للأوساط الحسابية من جدول (٣) ، حيث : $\bar{\bar{س}}$ = ٨,٢ سنة .

وفي الواقع فإن الوسط الحسابي للتوزيع العيني $\mu_{\bar{س}}$ ما هـو إلا الوسط الحسابي للأوساط الحسابية ، أي أن :

$$\mu_{\bar{س}} = \bar{\bar{س}} \qquad (٣)$$

وفي مثال (١) وجدنا أن الوسط الحسابي للمجتمع μ = ٨,٢ سنة أيضاً . وهذا التساوي بين الوسط الحسابي للتوزيع العيني ($\mu_{\bar{س}}$) والوسط الحسابي للمجتمع (μ) ليس وليد للصدفة ولكن هذا التساوي هو نتيجة لخاصية مهمة ألا وهي :

الوسط الحسابي للأوساط الحسابية للعينات يساوي الوسط الحسابي للمجتمع أي أنه

$$\mu_{\bar{س}} = \mu \qquad (٤)$$

ومن جدول (٦) يمكننا الحصول على تباين توزيع معاينة $\bar{س}$ ، والذي يرمز له بالرمز ($\sigma^{٢}_{\bar{س}}$) كما يلي :

$$\sigma^{٢}_{\bar{س}} = \frac{١}{مجـ ك} \left[مجـ \bar{س}^{٢} ك - \frac{(مجـ \bar{س} ك)^{٢}}{مجـ ك} \right]$$

$$= \frac{١}{٢٥} \left[١,٨٧٣ - \frac{(٢٠٥)^{٢}}{٢٥} \right] = ٧,٦٨$$

$$\therefore \sigma_{\bar{س}} = \sqrt{٧,٦٨} = ٢,٧٧١ \text{ سنة}$$

ويسمى الانحراف المعياري لتوزيع المعاينة بالخطأ المعياري standard error .

ومـن الملاحـظ أن الانحـراف المعيـاري للتوزيـع العينـي ($\sigma_{\bar{س}}$ = ٢,٧٧١ سنة) لا يساوي الانحراف المعياري للمجتمع الذي سبق وحصلنا على قيمته في مثال (١) حيث (σ = ٣,٩١٩ سنة) . ولكـن توجد علاقة بينهما ، ألا وهي :

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} \qquad (٥)$$

وبالفعل : $\sigma_{\bar{س}} = \frac{٣,٩١٩}{\sqrt{٢}} =$ ٢,٧٧١ سنة

ولقد سبق وحصلنا على هذه القيمة عند حساب $\sigma_{\bar{س}}$ من جـدول (٦) . وتكون هذه العلاقة بين الانحراف المعياري لتوزيع معاينـة $\bar{س}$ وبيـن الانحراف المعياري للمجتمع – علاقة (٥) – صحيحة إذا تـم سـحب العينات بإرجاع من مجتمع محدود الحجم finite population ، حيـث لا يؤثر سحب مفردة من المجتمع على سحب مفردة أخرى ، ومن ثم يكـون هذين الحدثين مستقلين إحصائيا ، وتكون العلاقة (٥) صحيحة أيضا فـي حالة المجتمعات اللانهائية الحجم – سواء كان السحب بإرجاع و بـدون إرجاع – حيث لا يتأثر حجم المجتمع بسحب مفردة منه ، ومن ثم يكـون حدث سحب مفردة مستقلا عن حدث سحب مفردة أخرى . ويتحقق هـذا الاستقلال الإحصائي عندما يكون حجم المجتمع م كبيرا بالنسـبة لحجـم العينة ن ، أي عندما $\frac{ن}{م} <$ ٠,٠٥ م[١].

---

(1) Kohler, H., "Statistics for Business & Economics", Harper Collins College Publisher, 1994, pp. 305, 306.

Mann P. S., "Statistics for Business & Economics", John Wiley & Sons., Inc., 1995, p. 371.

وخلاصة القول :

عندما يكون السحب بإرجاع ، أو عندما تكون النسبة بين حجم العينة ن وحجم المجتمع م هي : $\frac{ن}{م} < ٠,٠٥$ ، أي في حالة تحقق الاستقلال الإحصائي ، فإن :

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} \qquad (٦)$$

وفي أغلب الأحيان يكون حجم العينة صغيرا نسبيا بالنسبة لحجم المجتمع مما يجعل العلاقة (٦) أكثر استخداما في الحياة العملية .

٢ ــ حالة السحب بدون إرجاع :

في حالة السحب بدون إرجاع فإن أي مفردة في المجتمع لا يمكن أن تسحب إلا مرة واحدة لتكوين العينة . فإذا أخذنا جميع العينات الممكنة المكونة من مفردتين بدون إرجاع فسيكون لدينا ما يلي :

أ ب	أ حـ	أ د	أ هـ	ب حـ	ب د	ب هـ	حـ د	حـ هـ	د هـ
٤،٦	١٥،٦	٦،٦	١٠،٦	١٥،٤	٦،٤	١٠،٤	٦،١٥	١٠،١٥	١٠،٦

أي أن لدينا ١٠ عينات . ويمكن الحصول على عدد العينات الكلية الممكنة بدون إرجاع عن طريق التوافيق :

$$\text{عدد العينات الكلية الممكنة} = {}^{م}ق_{ن} = \frac{م(م-١)(م-٢)\dots(م-ن-١)}{ن(ن-١)(ن-٢)\dots \times ٣ \times ٢ \times ١}$$

$$= {}^{٥}ق_{٢} = \frac{٥ \times ٤}{٢ \times ١} = ١٠ \text{ عينات .}$$

ويبين جدول (٧) جميع العينات الممكنة المكونة من مفردتيـــن والوسط الحسابي لكل منها :

جدول (٧)

جميع العينات الممكنة المكونة من مفردتين في حالة السحب بدون إرجاع والوسط الحسابي لكل منها

العينة	مفردات العينة	الوسط الحسابي $\bar{س}$
أ ب	٦ ، ٤	٥
أ حـ	٦ ، ١٥	١٠,٥
أ د	٦ ، ٦	٦
أ هـ	٦ ، ١٠	٨
ب حـ	٤ ، ١٥	٩,٥
ب د	٤ ، ٦	٥
ب هـ	٤ ، ١٠	٧
حـ د	١٥ ، ٦	١٠,٥
حـ هـ	١٥ ، ١٠	١٢,٥
د هـ	٦ ، ١٠	٨
المجموع		٨٢

وحصلنا على الوسط الحسابي بقسمة مفردات العينة على عددها ، فمثـــلاً عندما كانت المفردات (٦ ، ٤) فإن :

$$\bar{س} = \frac{٦ + ٤}{٢} = ٥$$

ومن جدول (٧) يتضح لنا أن الأوساط الحسابية تختلف قيمتها من عينة إلى أخرى وفقا لمفردات العينة . كما أن كل من هذه الأوساط الحسابية يختلف عن الوسط الحسابي للمجتمع μ ، حيث μ = ٨,٢ سنة كما سبق وحصلنا عليها في مبحث (١ ــ ١) . ومن جدول (٧) يمكننا الحصول على التوزيع التكراري للأوساط الحسابية كما هو موضح في جدول (٨) .

جدول (٨)

التوزيع التكراري الأوساط الحسابية سَ

عندما تكون العينة مكونة من مفردتين ، في حالة السحب بدون إرجاع

الوسط الحسابي سَ	التكرار ك
٥	٢
٦	١
٧	١
٨	٢
٩,٥	١
١٠,٥	٢
١٢,٥	١
المجموع	١٠

وبقسمة التكرار على مجموع التكرارات نحصل على التكرار النسبي ، وهو يمثل الاحتمال في كل فئة . ويبين جدول (٩) توزيع المعاينة للأوساط الحسابية سَ عندما تكون العينة مكونة من مفردتين .

جدول (٩)

توزيع المعاينة للأوساط الحسابية $\bar{س}$

عندما تكون العينة مكونة من مفردتين

الوسط الحسابي $\bar{س}$	التكرار النسبي ح ($\bar{س}$)
٥	$\frac{٢}{١٠}$
٦	$\frac{١}{١٠}$
٧	$\frac{١}{١٠}$
٨	$\frac{٢}{١٠}$
٩,٥	$\frac{١}{١٠}$
١٠,٥	$\frac{٢}{١٠}$
١٢,٥	$\frac{١}{١٠}$
المجموع	١,٠٠

وكما سبق وذكرنا ، فإنه يرمز للوسط الحسابي لتوزيع معاينة $\bar{س}$ بالرمز $\mu_{\bar{س}}$ أو $\bar{\bar{س}}$. ويرمز للانحراف المعياري لتوزيع معاينة $\bar{س}$ بالرمز $\sigma_{\bar{س}}$.

يمكن الحصول على $\bar{\bar{س}}$ و $\sigma_{\bar{س}}$ باستخدام جدول (٧) ، أي حالة القيم غير المبوبة ، كما هو موضح من جدول (١٠) .

جدول (١٠)

حساب الوسط الحسابي والانحراف المعياري

للأوساط الحسابية في حالة السحب بدون إرجاع

الوسط الحسابي $\bar{س}$	$\bar{س} - \bar{\bar{س}}$	$(\bar{س} - \bar{\bar{س}})^2$
٥	-٣,٢	١٠,٢٤
١٠,٥	٢,٣	٥,٢٩
٦	-٢.٢	٤,٨٤
٨	-٠,٢	٠,٠٤
٩,٥	١,٣	١,٩٩
٥	-٣,٢	١٠,٢٤
٧	-١,٢	١,٤٤
١٠,٥	٢,٣	٥,٢٩
١٢,٥	٤,٣	١٨,٤٩
٨	-٠,٢	٠,٠٤
٨٢	صفر	٥٧,٦٠

$$\mu_{\bar{س}} = \bar{\bar{س}} = \frac{مج\ \bar{س}}{عدد\ العينات} = \frac{٨٢}{١٠} = ٨,٢ \text{ سنة}$$

$$\sigma^2_{\bar{س}} = \frac{مج\ (\bar{س} - \bar{\bar{س}})^2}{عدد\ العينات} = \frac{٥٧,٦٠}{١٠} = ٥,٧٦$$

$$\sigma_{\bar{س}} = \sqrt{٥,٧٦} = ٢,٤ \text{ سنة}$$

كما يمكن الحصول على $\bar{\bar{س}}$ و $\sigma_{\bar{س}}$ باستخدام جـــدول (٨) ، أي حالة القيم المبوبة ، كما هو موضح من جدول (١١) .

جدول (١١)

حساب الوسط الحسابي والانحراف المعياري

لتوزيع معاينة $\bar{س}$ في حالة السحب بدون إرجاع

الوسط الحسابي $\bar{س}$	التكرار ك	$\bar{س}$ ك	$\bar{س}^{٢}$ ك
٥	٢	١٠	٥٠
٦	١	٦	٣٦
٧	١	٧	٤٩
٨	٢	١٦	١٢٨
٩,٥	١	٩,٥	٩٠,٢٥
١٠,٥	٣	٢١	٢٢٠,٥٠
١٢,٥	١	[illegible]	١٥٦,٢٥
المجموع	١٠	٨٢	٧٣٠

$$\mu_{\bar{س}} = \frac{\text{مجـ } \bar{س} \text{ ك}}{\text{مجـ ك}} = \frac{[illegible]}{١٠} = ٨,٣ \text{ سنة}$$

وكما سبق أن وجدنا في حالة السحب مع إرجاع ، فإن الوسط الحسابي للتوزيع العيني في حالة السحب بدون إرجاع يساوي الوسط الحسابي للمجتمع . أي أن :

$$\mu_{\bar{س}} = \mu \qquad (٧)$$

أما تباين توزيع معاينة $\bar{س}$ في حالة السحب بدون إرجاع ،

$$\sigma^2_{\bar{س}} = \frac{1}{مجـ ك} \left[مجـ \bar{س}^2 ك - \frac{(مجـ \bar{س} ك)^2}{مجـ ك} \right]$$

$$= \frac{1}{١٠} \left[٧٣٠ - \frac{(٨٢)^2}{١٠} \right] = ٥,٧٦$$

$$\sigma_{\bar{س}} = \sqrt{٥,٧٦} = ٢,٤ \text{ سنة}$$

وبوجه عام ، عندما يتم السحب بدون إرجاع ، فإن سحب مفردة من المجتمع لتكوين العينة يؤثر على سحب أي مفردة أخرى وهنا ، يكون الحدثين غير مستقلين . ويتحقق عدم الاستقلال الإحصائي أيضا عندما يكون حجم المجتمع صغير نسبيا بالنسبة لحجم العينة أي عندما يتحقق الشرط : $\frac{ن}{م} \geq ٠,٠٥$.[1]

عندما يتم السحب بدون إرجاع ، أو عندما تكون النسبة بين حجم العينة ن وحجم المجتمع م هي : $\frac{ن}{م} \geq ٠,٠٥$ ، أي في حالة عدم تحقق الاستقلال الإحصائي ، فإن : $$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} \sqrt{\frac{م - ن}{م - ١}} \qquad (٨)$$

ويسمى المقدار $\frac{م - ن}{م - ١}$ بمعامل التصحيح correction factor ولقد سبق ووجدنا من جدولي (١) ، (٢) أن : $\sigma = ٣,٩١٩$

$$\sigma_{\bar{س}} = \frac{٣,٩١٩}{\sqrt{٢}} \sqrt{\frac{٥ - ٢}{٥ - ١}} = ٢,٣٩٩٨٨ \approx ٢,٤ \text{ سنة} .$$

(1) Kohler,H., op.cit. Pp. 305, 306

مثال (٣) :

إذا كان حجم المجتمع ٦٠٠٠ مفردة ، وكـان الوسـط الحسـابي والانحراف المعياري في هذا المجتمع هما : ٢٦ ، ٣ علـى التوالـي ، فالمطلوب حساب كل من الوسط الحسابي والانحراف المعياري لتوزيـع معاينة $\bar{س}$ ، إذا كان حجم العينة :

أولاً : ٣٠ مفردة

ثانياً : ٥٠٠ مفردة .

الحـل :

أولاً : ن = ٣٠ ، م = ٦٠٠٠ ، μ = ٢٦ ، σ = ٣

$\mu_{\bar{س}}$ = ؟ ، $\sigma_{\bar{س}}$ = ؟

الوسط الحسابي لتوزيع معاينة $\bar{س}$:

$$\mu_{\bar{س}} = \mu = ٢٦$$

في هذه الحالة $\frac{ن}{م} = \frac{٣٠}{٦٠٠٠} = ٠,٠٠٥ < ٠,٠٥$

ومن ثم فإن الانحراف المعياري لتوزيع معاينة $\bar{س}$ نحصل عليه مـن القانون :

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} = \frac{٣}{\sqrt{٣٠}} = ٠,٥٤٨$$

ثانياً : ن = ٥٠٠ ، م = ٦٠٠٠ ، μ = ٢٦ ، σ = ٣

$\mu_{\bar{س}}$ = ؟ ، $\sigma_{\bar{س}}$ = ؟

الوسط الحسابي لتوزيع معاينة $\bar{س}$:

$$\mu_{\bar{س}} = \mu = ٢٦$$

في هذه الحالة $\frac{ن}{م} = \frac{٥٠٠}{٦٠٠٠} = ٠,٠٨٣ > ٠,٠٥$

ومن ثم فإن الانحراف المعياري لتوزيع معاينة $\bar{س}$ نحصل عليه من القانون :

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} \sqrt{\frac{م - ن}{م - ١}}$$

$$= \frac{٣}{\sqrt{٥٠٠}} \sqrt{\frac{٦٠٠٠ - ٥٠٠}{٦٠٠٠ - ١}}$$

$$= \frac{٣}{\sqrt{٥٠٠}} (٠,٩٥٧٥) = ٠,١٢٨$$

(١ ــ ٣) الوسط الحسبي والانحراف المعياري لتوزيع معاينة $\bar{س}$:

لقد عرفنا في بداية هذا الفصل العينة العشوائية البسيطة حيث تكون لكل مفردة نفس الفرصة في تكوين العينة . ومن ثم فإن كل مفردة من مفردات العينة تكون مستقلة عن باقي المفردات .

نظرية ١ :

العينة العشوائية البسيطة هي تلك العينة التي تكون كل مشاهداتها ($س_١$ ، $س_٢$ ، . . . . ، $س_ن$) مستقلة . ويكون توزيع كل مشاهدة س هو نفسه توزيع المجتمع الذي سحبت منه ، أي أن :

توزيع $س_١$ = توزيع $س_٢$ = . . . = توزيع $س_ن$ = توزيع المجتمع

ومن ثم فإن كل مشاهدة في العينة يكون :

وسطها الحسابي = الوسط الحسابي للمجتمع الذي سحبت منه

وانحرافها المعياري = الانحراف المعياري للمجتمع الذي سحبت منه

وبدراسة المتغيرات العشوائية فإن :

ت (س + ص) = ت (س) + ت (ص) (٩)

ت (أ س + ب ص) = أ ت (س) + ب ت (ص) (١٠)

وإذا كانت س ، ص مستقلتان فإن :

تباين (س + ص) = تباين س + تباين ص (١١)

تباين (أ س + ب ص) = $أ^٢$ تباين س + $ب^٢$ تباين ص (١٢)

ولقد بينا في المبحث السابق أن الوسط الحسابي لتوزيع معاينة $\overline{س}$ يساوي الوسط الحسابي للمجتمع ، وفيما يلي برهان هذه المعادلة .

إذا كان لدينا المشاهدات الآتية :

$س_١$ ، $س_٢$ ، . . . . . ، $س_ن$

فبالتعريف ، الوسط الحسابي لهذه القيم هو :

$$\overline{س} = \frac{1}{ن} (س_{١} + س_{٢} + . . . + س_{ن})$$

$$= \frac{1}{ن} (س_{١}) + \frac{1}{ن} (س_{٢}) + . . . + \frac{1}{ن} (س_{ن})$$

فإذا كانت كل من : $س_{١}$ ، $س_{٢}$ ، . . . . ، $س_{ن}$ متغيرات عشوائية فإن $\overline{س}$ بدوره يكون متغيرا عشوائيا . وباستخدام معادلة (١٠) ، فإن

$$ت (\overline{س}) = \frac{1}{ن} ت (س_{١}) + \frac{1}{ن} ت (س_{٢}) + ... + \frac{1}{ن} ت (س_{ن})$$

$$\therefore ت (\overline{س}) = \frac{1}{ن} [ت (س_{١}) + ت (س_{٢}) + ... + ت (س_{ن})] \quad (١٣)$$

وباستخدام نظرية ١ ، فإن كل مشاهدة (س) لها نفس توزيع المجتمع الذي سحبت منه ووسطها الحسابي يساوي الوسط الحسابي للمجتمع μ . أي أن :

$$ت (س_{١}) = ت (س_{٢}) = . . . = ت (س_{ن}) = \mu \quad (١٤)$$

ومن ثم ، فإن معادلة (١٣) تصبح :

$$ت (\overline{س}) = \frac{1}{ن} [\underbrace{\mu + \mu + ... + \mu}_{\text{ن مرة}}]$$

$$= \frac{1}{ن} (ن \mu) = \mu$$

أي أن :

$$\boxed{ت (\overline{س}) = \mu} \quad (١٥)$$

وهذا معناه أن الوسط الحسابي لتوزيع معاينة $\overline{س}$ يساوي الوسط الحسابي للمجتمع . وبما أن الوسط الجسابي $\overline{س}$ متغيرا عشوائيا قيمته تختلف من عينة إلى أخرى ، فتارة تزيد عن قيمة μ وتارة تنخفض عن قيمة μ ، مما يدفعنا إلى معرفة تباين توزيع معاينة $\overline{س}$.

ولقد سبق وبينا أن :

$$\overline{س} = \frac{١}{ن} \left(س_{١} + س_{٢} + \ldots + س_{ن} \right)$$

أي أن :

$$\overline{س} = \frac{١}{ن} \left(س_{١} \right) + \frac{١}{ن} \left(س_{٢} \right) + \ldots + \frac{١}{ن} \left(س_{ن} \right)$$

وبما أن $س_{١}$ ، $س_{٢}$ ، . . . ، $س_{ن}$ متغيرات عشوائية مستقلة ، وباستخدام معادلة (١١) :

$$\text{تباين } \overline{س} = \text{تباين } \left(\frac{١}{ن} س_{١} \right) + \text{تباين } \left(\frac{١}{ن} س_{٢} \right) + \ldots + \text{تباين } \left(\frac{١}{ن} س_{ن} \right)$$

وبما أن $س_{١}$ ، $س_{٢}$ ، . . . ، $س_{ن}$ مستقلة ، وباستخدام قانون (١٢) ، فإن :

$$\text{تباين } \overline{س} = \left(\frac{١}{ن} \right)^{٢} \text{تباين } س_{١} + \left(\frac{١}{ن} \right)^{٢} \text{تباين } س_{٢} + \ldots + \left(\frac{١}{ن} \right)^{٢} \text{تباين } س_{ن} \qquad (١٦)$$

ووفقا لنظرية ١ ، فإن الانحراف المعياري لكل مشاهدة يساوي الانحراف المعياري للمجتمع الذي سحبت منه ، أي أن :

$$\text{انحراف معياري } س_{١} = \text{انحراف معياري } س_{٢} = \cdots = \text{انحراف معياري } س_{ن}$$

$$= \sigma$$

وهذا يعني أن :

$$\text{تباين } س_{١} = \text{تباين } س_{٢} = \cdots = \text{تباين } س_{ن} = \sigma^{٢} \qquad (١٧)$$

وباستخدام معادلة (١٧) ، تصبح معادلة (١٦) كما يلي :

$$\text{تباين } \bar{س} = \left(\frac{1}{ن}\right)^2 \sigma^2 + \left(\frac{1}{ن}\right)^2 \sigma^2 + \cdots + \left(\frac{1}{ن}\right)^2 \sigma^2$$

$$= \frac{1}{ن^2} \left(\underbrace{\sigma^2 + \sigma^2 + \cdots + \sigma^2}_{\text{ن مرة}} \right)$$

أي أن : $\text{تباين } \bar{س} = \frac{1}{ن^2} \left(ن \ \sigma^2 \right)$

$$\text{تباين } \bar{س} = \frac{\sigma^2}{ن}$$

$$\therefore \quad \boxed{\sigma^2_{\bar{س}} = \frac{\sigma^2}{ن}} \qquad (١٨)$$

وبأخذ الجذر التربيعي للطرفين ، فإن الانحراف المعياري لتوزيع معاينة $\bar{س}$ (أي الخطأ المعياري) هو :

$$\boxed{\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}} \qquad (١٩)$$

(١ ــ ٤) شكل توزيع معاينة $\bar{س}$:

من نظرية ١ ، رأينا أن توزيع كل مشاهدة س ، في العينة العشوائية البسيطة ، لها نفس توزيع المجتمع الذي سحبت منه . والسؤال الذي يثار هنا ما هو شكل توزيع معاينة $\bar{س}$. وهنا يجب التفرقة بين حالتين : حالة العينات المسحوبة من مجتمع يتبع التوزيع المعتدل وحالة العينات المسحوبة من مجتمع لا يتبع التوزيع المعتدل .

١ ــ العينات المسحوبة من مجتمع يتبع التوزيع المعتدل :

نظرية ٢ :

إذا كانت : $س_١$ ، $س_٢$ ، . . . . ، $س_ن$ مفردات عينة عشوائية مكونة من ن مفردة ، ومسحوبة من مجتمع معتدل ع (μ ، $\sigma^٢$) ، فإن توزيع معاينة $\bar{س}$ يكون أيضا معتدلا ع (μ ، $\frac{\sigma^٢}{ن}$) .

وهذا يعني إذا المجتمع معتدل ومتوسطه μ وتباينه $\sigma^٢$ ، فإن توزيع معاينة الوسط الحسابي لجميع العينات الممكنة التي لها نفس الحجم ، والمسحوبة من هذا المجتمع ، يكون له الخصائص التالية :

١ ــ هذا التوزيع معتدل .

٢ ــ الوسط الحسابي لهذا التوزيع يساوي الوسط الحسابي للمجتمع

أي : $\mu_{\bar{س}} = \mu$

٣ ــ تباين هذا التوزيع يساوي تباين المجتمع مقسوما على حجم العينة

أي : $\sigma^٢_{\bar{س}} = \frac{\sigma^٢}{ن}$

وبأخذ الجذر التربيعي للطرفين ، فإن الانحراف المعياري لتوزيع المعاينة (المسمى بالخطأ المعياري) يساوي الانحراف المعياري للمجتمع مقسوما على الجذر التربيعي لحجم العينة :

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$$

ومن دراستنا للتوزيع المعتدل ع (μ ، σ^2) ، رأينا أنه يمكن حساب الاحتمالات المختلفة عن طريق إيجاد المساحات المناظرة تحت المنحنى المعتدل المعياري ع (صفر ، ١) ، أي المنحنى المعتدل الذي يكون وسطه الحسابي صفر وانحرافه المعياري ١ . لذلك نحول قيم المتغير العشوائي س إلى قيم معيارية Z ، حيث :

$$Z = \frac{س - \mu}{\sigma}$$

ثم نستخدم جداول المنحنى المعتدل المعياري لإيجاد المساحات المطلوبة .

إذا كان لدينا عينة عشوائية مكونة من ن مفردة ، وكان توزيع معاينة وسطها الحسابي معتدلا ع (μ ، $\frac{\sigma^2}{ن}$) ، وكان المطلوب إيجاد الاحتمالات المختلفة تحت المنحنى المعتدل ، فإننا نقوم بتحويل قيم س إلى قيم معيارية Z حيث :

$$Z = \frac{\bar{س} - \mu}{\frac{\sigma}{\sqrt{ن}}}$$

ثم نستخدم جداول المنحنى المعتدل المعياري لإيجاد المساحات المطلوبة .

مثال (٤) :

في أحد امتحانات مادة الإحصاء في السنة الثانية لكلية التجارة ، كانت درجات الطلبة تتوزع توزيعاً معتدلاً ، وسطه الحسابي ٧٠ درجة وانحرافه المعياري ١٠ درجات .

أولا : إذا سحبت ورقة امتحان واحدة عشوائيا ، أوجد احتمال أن يكـــون هذا الطالب حاصل على درجة أكبر من ٧٥ .

ثانيا : إذا سحبت عينة عشوائية من ٩ طلاب ، احسب الوسط الحســـابي والخطأ المعياري لتوزيع معاينة $\overline{س}$ ، ثم أوجد احتمال أن يكـــون الوسط الحسابي لهذه العينة أكبر من ٧٥ درجة .

ثالثا : إذا سحبت عينة عشوائية من ٤٩ طالب ، احسب الوسط الحســابي والخطأ المعياري لتوزيع معاينة $\overline{س}$ ، ثم أوجد احتمال أن يكـــون الوسط الحسابي لهذه العينة ينحصر بين ٦٨ درجة و ٧٢ درجة .

الحـــــل :

μ = ٧٠ درجة ، σ = ١٠ درجات

أولا : بما أن درجات الطلبة تتوزع توزيعا معتدلا ، فإن درجة الطـــالب المسحوبة من هذا المجتمع تتوزع هي أيضا توزيعا معتدلا .

$$ح\,(س > ٧٥) = ح\left(\frac{س - \mu}{\sigma} > \frac{٧٥ - ٧٠}{١٠}\right)$$

$$= ح\,(Z > ٠,٥)$$

$$= ١ - \Phi\,(٠,٥)$$

$$= ١ - ٠,٦٩١٥$$

$$= ٠,٣٠٨٥$$

ثانيا : ن = ٩ ، μ = ٧٠ درجة ، σ = ١٠ درجات .

الوسط الحسابي لتوزيع معاينة $\overline{س}$:

$\mu_{\bar{س}} = \mu = ٧٠$ درجة

الخطأ المعياري لتوزيع معاينة $\bar{س}$:

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} = \frac{١٠}{\sqrt{٩}} = \frac{١٠}{٣} = ٣,٣٣ \text{ درجة}$$

$$ح(\bar{س} > ٧٥) = ح\left(Z > \frac{٧٥ - ٧٠}{٣,٣٣}\right)$$

$$= ح(Z > ١,٥)$$

$$= ١ - \Phi(١,٥)$$

$$= ١ - ٠,٩٣٣٢$$

$$= ٠,٠٦٦٨$$

ثالثا : ن = ٤٩ ، μ = ٧٠ درجة ، σ = ١٠ درجات .

الوسط الحسابي لتوزيع معاينة $\bar{س}$:

$\mu_{\bar{س}} = \mu = ٧٠$ درجة

الخطأ المعياري لتوزيع معاينة $\bar{س}$:

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} = \frac{١٠}{\sqrt{٤٩}} = \frac{١٠}{٧} = ١,٤٣ \text{ درجة}$$

$$ح(٦٨ \le \bar{س} \le ٧٥) = ح\left(\frac{٦٨ - ٧٠}{١,٤٣} \le Z \le \frac{٧٢ - ٧٠}{١,٤٣}\right)$$

$$= ح(-١,٤٠ \le Z \le ١,٤٠)$$

$$= \Phi(١,٤) - \Phi(-١,٤)$$

$$= \Phi(١,٤) - [١ - \Phi(١,٤)]$$

$$= ٢\Phi(١,٤) - ١$$

$$= ٢(٠,٩١٩٢) - ١$$

$$= ٠,٨٣٨٤$$

٢ ـ العينات المسحوبة من مجتمع لا يتبع التوزيع المعتدل :

قد يحدث في كثير من الأحيان أن يكون المجتمع الذي تسحب منه العينات ليس معتدلا . فقد يكون ملتويا نحو اليمين أو نحو اليسار . في هذه الحالة نطبق نظرية النهاية المركزية Central limit theorem .

نظرية ٣ :

نظرية النهاية المركزية :

إذا كانت : $س_١$ ، $س_٢$ ، . . . ، $س_ن$ مفردات عينة عشوائية مكونة من ن مفردة ، ومسحوبة من مجتمع لا يتبع التوزيع المعتدل ، ووسطه الحسابي μ وتباينه $\sigma^٢$ ، فإن توزيع معاينة $\bar{س}$ يقترب من التوزيع المعتدل ع (μ ، $\frac{\sigma^٢}{ن}$) كلما ازداد حجم العينة (ن) .

تنطبق نظرية النهاية المركزية على المجتمعات الكبيرة فقط ، أي عندما ن $\geq$ ٣٠ .

وفقا لنظرية النهاية المركزية ، فإذا كان توزيع المجتمع التي تسحب فيه العينات ليس معتدلا ، فإن شكل توزيع معاينة $\bar{س}$ لا يكون معتدلا بالضبط ، ولكنه يكون قريبا من الاعتدال عندما يكون حجم العينة كبيرا (ن $\geq$ ٣٠) .

ومن الملاحظ أن : $\mu_{\bar{س}} = \mu$

وأن : $\sigma^٢_{\bar{س}} = \frac{\sigma^٢}{ن}$

أي : $\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$

ولكن يجب أن نتذكر أنه لكي تكون : $\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$ يجب تحقيق الشرط القائل بأن نسبة حجم العينة إلى حجم المجتمع : $\frac{ن}{م} < ٠,٠٥$ ؛ أما إذا كانت $\frac{ن}{م} \geq ٠,٠٥$ فيجب استخدام معامل التصحيح عند حساب $\sigma_{\bar{س}}$ كما سبق وذكرنا في معادلة (٨) .

مثال (٥) :

في أحد المصانع كان الوسط الحسابي لأجور العمال ٣٠٠ جنيه وانحرافه المعياري ٥٠ جنيه . ولقد سحبت عينة عشوائية من ٤٠ عامل من عمال هذا المصنع ، أوجد :

أولاً : احتمال أن يكون الوسط الحسابي للعينة أقل من ٢٧٥ جنيه .

ثانياً : احتمال أن يكون الوسط الحسابي للعينة أكبر من ٣٢٠ جنيه .

ثالثاً : احتمال أن يكون الوسط الحسابي للعينة بين ٢٨٠ و ٣١٥ جنيه .

الحل :

هنا $\mu = ٣٠٠٠$ جنيه ، $\sigma = ٥٠$ جنيه ، $ن = ٤٠$

وحيث أن توزيع مجتمع أجور العمال غير معطى ، أي غير معلوم ، ونظراً لأن حجم العينة كبير (ن > ٣٠) ، إذاً يمكننا تطبيق نظرية النهاية المركزية . ومن ثم فإن توزيع معاينة $\bar{س}$ يكون معتدلاً تقريباً بوسط حسابي : $\mu_{\bar{س}} = \mu = ٣٠٠٠$ جنيه

وخطأ معياري : $\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} = \frac{٥٠}{\sqrt{٤٠}} = ٧,٩١$ جنيه

<u>أولا</u> : ح ($\bar{س}$ > ٢٧٥) = ح ($\frac{\bar{س} - \mu}{\sigma_{\bar{س}}}$ > $\frac{٢٧٥ - ٣٠٠}{٧,٩١}$)

= ح (Z > -٣,١٦)

= ١ - Φ (٣,١٦)

= ١ - ٠,٩٩٢

= ٠,٠٠٠٨

<u>ثانيا</u> : ح ($\bar{س}$ < ٣٢٠) = ح ($\frac{\bar{س} - \mu}{\sigma_{\bar{س}}}$ < $\frac{٣٢٠ - ٣٠٠}{٧,٩١}$)

= ح (Z < ٢,٥٣)

= ١ - Φ (٢,٥٣)

= ١ - ٠,٩٩٤٣

= ٠,٠٠٥٧

<u>ثالثا</u> :

ح (٢٨٠ ≤ $\bar{س}$ ≤ ٣١٥) = ح ($\frac{٢٨٠ - ٣٠٠}{٧,٩١}$ ≤ Z ≤ $\frac{٣١٥ - ٣٠٠}{٧,٩١}$)

= ح (- ٢,٥٣ ≤ Z ≤ ١,٩٠)

= Φ (١,٩٠) - Φ (- ٢,٥٣)

= Φ (١,٩٠) - [١ - Φ (٢,٥٣)]

= ٠,٩٧٣١ - (١ - ٠,٩٩٤٣)

= ٠,٩٧١٣ - ١ + ٠,٩٩٤٣

= ٠,٩٦٥٦

(١ ـــ ٥) توزيع معاينة النسبة ق :

في كثير من الأحيان يتم تقسيم المجتمع إلى نوعين من المفردات : المفردات التي تتصف بصفة معينة والمفردات التي لا تتصف بهذه الصفـــة . فمثلا قد ينقسم المجتمع إلى مدخن وغير مدخن ، أو إلى إناث وذكـــور ، وقد ينقسم مجتمع إنتاج مصنع معين إلى إنتاج معيب وإنتاج غير معيـب . فإذا سحبت مفردة من أحد هذه المجتمعات بطريقة عشوائية ، فإن :

احتمال تمتع هذه المفردة بصفة معينة = نسبة هذه الظاهرة في المجتمع = θ

أما احتمال عدم تمتع المفردة المسحوبة بهذه الصفـــة = ١ − θ . فمثلا إذا كانت نسبة المدخنين في مجتمع معين هي θ = ٠,٤ فإن نســبة غير المدخنين في هذا المجتمع تساوي (١ − θ) = ١ − ٠,٤ = ٠,٦

وإذا سحبنا عينة عشوائية حجمها ن من هذا المجتمع ، وإذا كـــان س هو المتغير العشوائي الذي يمثل عدد الذين يتصفون بهذه الصفة فـــي المجتمع ، فإن س تكون متغيرا عشوائيا له توزيع ذي الحديـــن . ويمكـــن حساب النسبة (ق) في العينة بقسمة عدد المفردات التي تتمتع بالصفـــة المعينة على حجم العينة ، أي أن :

$$\text{ق} = \frac{\text{عدد المفردات التي تتمتع بالصفة}}{\text{حجم العينة}} = \frac{\text{س}}{\text{ن}} \qquad (٢٠)$$

وبما أن عدد المفردات التي تتمتع بالصفة تختلف من عينة إلى أخرى، إذن فالنسبة ق هي متغير عشوائي له توزيع احتمالي هو توزيع معاينة نســبة العينة ق . ويوضح مثال (٦) هذا التوزيع .

مثال (٦) :

في أحد فصول الدراسات العليا كان عدد الطلبة المتقدمين لامتحان الإحصاء ٥ طلبة . وكانت نتيجة الامتحان كما يلي :

رقم جلوس الطالب	١	٢	٣	٤	٥
النتيجة	ناجح	راسب	ناجح	ناجح	راسب

ويمكننا حساب نسبة النجاح في هذا المجتمع كما يلي :

$$\theta = \frac{٣}{٥} = ٠,٦٠$$

وإذا أخذنا جميع العينات الممكنة المكونة من مفردتين ، والتي يمكن سحبها من هذا المجتمع بدون إرجاع ، فإن :

عدد العينات الكلية الممكنة = $ق^{٥}_{٢}$ = $\frac{٥ \times ٤}{٢ \times ١}$ = ١٠ عينات

ويبين جدول (١٢) جميع العينات الممكنة في هذه الحالة ، والنسبة فـــي كل منها .

جدول (١٢)

جميع العينات الممكنة ونسبة كل منها

العينة	**مفردات العينة**	**نسبة العينة ق**
١ ، ٢	ناجح ، راسب	٠.٥
١ ، ٣	ناجح ، ناجح	١
١ ، ٤	ناجح ، ناجح	١
١ ، ٥	ناجح ، راسب	٠,٥
٢ ، ٣	راسب ، ناجح	٠,٥
٢ ، ٤	راسب ، ناجح	٠,٥
٢ ، ٥	راسب ، راسب	صفر
٣ ، ٤	ناجح ، ناجح	١
٣ ، ٥	ناجح ، راسب	٠,٥
٤ ، ٥	ناجح ، راسب	٠,٥

من هذا الجدول يمكننا اشتقاق جدول التوزيع التكراري للنسبة ق .

جدول (١٣)

التوزيع التكراري للنسبة ق

نسبة العينة ق	التكرار
صفر	١
٠,٥	٦
١	٣
المجموع	١٠

ومن هذا الجدول يمكننا الحصول على التكرار النسبي بقسمة التكرار على مجموع التكرارات ، وبذلك نحصل على التوزيع الاحتمـــالي للنســـبة ق . وهو يمثل توزيع معاينة النسبة ق . وهو المبين في جدول (١٤)

جدول (١٤)

توزيع معاينة النسبة ق

نسبة العينة (ق)	التكرار النسبي ح (ق)	ق ح (ق)	$ق^٢$ ح (ق)
صفر	$\frac{١}{١٠}$ = ٠,١	صفر	صفر
٠,٥	$\frac{٦}{١٠}$ = ٠,٦	٠,٣	٠,١٥
١	$\frac{٣}{١٠}$ = ٠,٣	٠,٣	٠,٣٠
المجموع	١	٠,٦	٠,٤٥

ومن هذا الجدول يمكننا حساب الوسط الحسابي وتباين توزيع معاينـــة ق كما يلي :

$\mu_{ق}$ = مجـ ق ح (ق) = ٠,٦

$\sigma^{2}_{ق}$ = مجـ ق^{2} ح (ق) − $\mu^{2}_{ق}$

= ٠,٤٥ − (٠,٦)^{2} = ٠,٠٩

ويتضح مما سبق أن $\mu_{ق}$ = θ = ٠,٦

وعلى وجه العموم ، يمكننا القول أن :

الوسط الحسابي لتوزيع معاينة النسبة ق يساوي نسبة المجتمع θ ، أي أن :

$$\mu_{ق} = \theta \qquad (٢١)$$

كما أن تباين توزيع معاينة ق يمكن الحصول عليه من القانون :

$$\sigma^{2}_{ق} = \frac{\theta\,(\,١ - \theta\,)}{ن} \qquad (٢٢)$$

ويكون الخطأ المعياري لتوزيع معاينة ق :

$$\sigma_{ق} = \sqrt{\frac{\theta\,(\,١ - \theta\,)}{ن}} \qquad (٢٣)$$

ويستخدم قانوني (٢٢) ، (٢٣) عندما تكون نسبة حجم العينة ن إلى حجم المجتمع م هي : $\frac{ن}{م} < ٠,٠٥$

أما إذا كانت نسبة حجم العينة ن إلى حجم المجتمع م :

$\frac{ن}{م} \geq ٠,٠٥$ ، فيجب استخدام معامل التصحيح ، أي أن :

عندما $\frac{ن}{م} \geq ٠.٠٥$ فإن :

$$\sigma_{ق} = \sqrt{\frac{\theta(1-\theta)}{ن}}\ \sqrt{\frac{م-ن}{م-1}} \qquad (24)$$

وفي أغلب الأحوال فإن حجم العينة يكون صغيرا بالنسبة لحجم المجتمع ومن ثم فإن قانون (٢٣) هو المستخدم .

وكما سبق وبينا بالنسبة لتوزيع معاينة $\bar{س}$ ، فإن نظرية النهاية المركزية تنطبق على توزيع معاينة ق .

نظرية ٤ :

تطبيقا لنظرية النهاية المركزية ، فعندما يكون حجم العينة كبيرا ، فإن توزيع معاينة النسبة ق يكون معتدلا تقريبا ع $\left(\theta ، \frac{\theta(1-\theta)}{ن}\right)$.

ويعتبر حجم العينة كبيرا إذا تحقق الشرطان :

$$ن\theta \geq 5$$

$$ن(1-\theta) \geq 5$$

مثال (٧) :

في أحد المجتمعات كانت نسبة المدخنين ٠,٣٥ ، فإذا سحبت عينة عشوائية من ١٠٠ مفردة فالمطلوب :

أولا : ما هو احتمال أن تكون نسبة المدخنين في هذه العينة أكبر من ٠,٤ ؟

ثانيا : ما هو احتمال أن تكون نسبة المدخنين في العينة في حدود ٠,٠٥ من نسبة المجتمع ؟

الحـــل :

<u>أولا</u> : $\theta = ٠,٣٥$ ، $(١ - \theta) = ٠,٦٥$ ، ق $= ٠,٤٠$

يمكن تقريب توزيع معاينة النسبة ق إلى التوزيع المعتدل إذا تحقق

الشرطان : ن $\theta \geq ٥$ ن $(١ - \theta) \geq ٥$

$$ن\,\theta = ١٠٠ \times ٠,٣٥ = ٣٥ > ٥$$

$$ن\,(١ - \theta) = ١٠٠ \times ٠,٦٥ = ٦٥ > ٥$$

∴ يمكن تقريب توزيع معاينة النسبة ق إلى التوزيع المعتدل .

حيث : $\mu_{ق} = \theta = ٠,٣٥$

$$\sigma_{ق} = \sqrt{\frac{\theta(١ - \theta)}{ن}} = \sqrt{\frac{٠,٣٥ \times ٠,٦٥}{١٠٠}}$$

$$= ٠,٠٤٨$$

$$ح\,(ق > ٠,٤٠) = ح\left(\frac{ق - \theta}{\sqrt{\frac{\theta(١ - \theta)}{ن}}} > \frac{٠,٤٠ - ٠,٣٥}{٠,٠٤٨}\right)$$

$$= ح\left(Z > \frac{٠,٠٥}{٠,٠٤٨}\right)$$

$$= ح\,(Z > ١,٠٤)$$

$$= ١ - \Phi(١,٠٤)$$

$$= ١ - ٠,٨٥٠٨$$

$$= ٠,١٤٩٢$$

ثانيا : المطلوب أن تكون نسبة المدخنين في العينة في حدود ٠,٠٥ مـــن نسبة المجتمع ، أي : $\theta \pm ٠,٠٥$

$٠,٣٥ \pm ٠,٠٥$

أي أن نسبة العينة تكون بين : ٠,٣٠ ، ٠,٤٠

$$ح\left(٠,٣٠ \le ق \le ٠,٤٠\right)$$

$$= ح\left(\frac{٠,٣٠ - ٠,٣٥}{٠,٠٤٨} \le \frac{ق - \theta}{\sqrt{\frac{\theta(١ - \theta)}{ن}}} \le \frac{٠,٤٠ - ٠,٣٥}{٠,٠٤٨}\right)$$

$$= ح\left(\frac{-٠,٠٥}{٠,٠٤٨} \le Z \le \frac{٠,٠٥}{٠,٠٤٨}\right)$$

$$= ح\left(-١,٠٤ \le Z \le ١,٠٤\right)$$

$$= \Phi(١,٠٤) - \Phi(-١,٠٤)$$

$$= \Phi(١,٠٤) - [١ - \Phi(١,٠٤)]$$

$$= ٢\Phi(١,٠٤) - ١$$

$$= ٢(٠,٨٥٠٨) - ١$$

$$= ٠,٧٠١٦$$

تمارين (١)

١ ـ في إحدى الشركات الصغيرة يقوم بالعمل ٤ موظفين فقط . وفيما يلي الأجور الشهرية (بالجنيهات) لهؤلاء الموظفين :

٤٠٠ ، ٥٠٠ ، ٥٠٠ ، ٦٠٠

والمطلوب :

أولا : حساب الوسط الحسابي والانحراف المعياري لهذا المجتمع

ثانيا : (أ) سحب جميع العينات الممكنة ذات الحجم ٢ في حالة السحب بإرجاع ، ثم تكوين توزيع معاينة $\bar{س}$.

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا التوزيع .

ثالثا : (أ) سحب جميع العينات الممكنة ذات الحجم ٢ في حالة السحب بدون إرجاع ، ثم تكوين توزيع معاينة $\bar{س}$.

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا التوزيع .

٢ ـ فيما يلي مجتمع مكون من ٣ مفردات :

١٠ ، ١٥ ، ٢٧ ، ٣٠

والمطلوب :

أولا : حساب الوسط الحسابي والانحراف المعياري لهذا المجتمع .

ثانيا : (أ) سحب جميع العينات الممكنة ذات الحجم ٣ في حالة السحب بإرجاع ، ثم تكوين توزيع معاينة $\bar{س}$.

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا التوزيع .

ثالثا : (أ) سحب جميع العينات الممكنة ذات الحجم ٣ في حالة السحب بدون إرجاع ، ثم تكوين توزيع معاينة $\bar{س}$.

(ب) حساب الوسط الحسابي والانحراف المعياري لهذا التوزيع .

٣ ــ كان عدد الموظفين في إحدى الشركات ١٠٠٠ موظف ، وكان الوسط الحسابي μ لأجور هؤلاء الموظفين هو ٨٠٠ جنيها شهريا بانحراف معياري (σ) ١٠٠ جنيه . فإذا سحبت عينة عشوائية من هذا المجتمع وتم حساب الوسط الحسابي ($\bar{س}$) لأجر الموظف منها ، فالمطلوب حساب الوسط الحسابي لتوزيع معاينة $\bar{س}$ ($\mu_{\bar{س}}$) والانحراف المعياري لهذا التوزيع ($\sigma_{\bar{س}}$) ، إذا كان حجم هذه العينة :

(أ) ٣٥ مفردة .

(ب) ٨٠ مفردة .

(حـ) ٣٠٠ مفردة .

٤ ــ في أحد المجتمعات كان الوسط الحسابي μ = ١٢٠ ، والانحراف المعياري σ = ١٥ .

أولا : إذا تم سحب عينة عشوائية من هذا المجتمع ، وكان الوسط الحسابي لتوزيع معاينة $\bar{س}$ هو : $\mu_{\bar{س}}$ = ١٢٠ ، والانحراف المعياري لهذا التوزيع هو : $\sigma_{\bar{س}}$ = ٣ . فإذا كانت العلاقة بين حجم العينة ن وحجم المجتمع م هي : $\frac{ن}{م}$ < ٠,٠٥ ، فما هو حجم العينة ؟

ثانياً : إذا تم سحب عينة عشوائية من هذا المجتمع وكان الوسط الحسابي لتوزيع معاينة $\bar{س}$ هو : $\mu_{\bar{س}}$ = ١٢٠ ، والانحراف المعياري لهذا التوزيع هو : $\sigma_{\bar{س}}$ = ٢ . فإذا كانت العلاقة بين حجم العينة ن وحجم المجتمع م هي : $\frac{ن}{م}$ < ٠,٠٥ ، فما هو حجم العينة ؟

٥ ـ كان توزيع سرعة السيارات المسافرة على إحدى الطرق السريعة معتدلاً بوسط حسابي ٩٠ كم / الساعة وانحرافه المعياري ١٠ كم / الساعة . ولقد تم سحب عينة عشوائية من ٢٥ سيارة مسافرة على هذا الطريق ، فإذا كان $\bar{س}$ هو الوسط الحسابي لسرعة السيارات في هذه العينة ، فالمطلوب :

أولاً : حساب الوسط الحسابي والانحراف المعياري لتوزيع معاينة $\bar{س}$.

ثانياً : ما هو شكل توزيع معاينة $\bar{س}$ ؟ ولماذا ؟

٦ ـ كانت الطرود الواردة لأحد مكاتب البريد لها توزيع ملتوي ناحية اليمين ، وكان وسطها الحسابي ٢,٥ كجم بانحراف معياري ٠,٩ كجم . وسحبت عينة عشوائية من ٤٠ طرد وارد لهذا المكتب . فإذا كان الوسط الحسابي لهذه العينة هو $\bar{س}$. فالمطلوب :

أولاً : حساب الوسط الحسابي والانحراف المعياري لتوزيع معاينة $\bar{س}$.

ثانياً : ما هو شكل توزيع معاينة $\bar{س}$ ؟ ولماذا ؟

٧ ــ في أحد البنوك كانت أرصدة الحسابات الجارية ملتوية ناحية اليميــن ، وكان وسطها الحسابي ٥٠٠ جنيها وانحراف معياري ٢٠٠ جنيـها . فإذا سحبت عينة عشوائية من ٥٠ حساب جاري من هـذا البنـك ، وإذا كانت $\bar{س}$ هي الوسط الحسابي للأرصدة في هذه العينة . فالمطلوب :

أولا : حساب الوسط الحســابي والانحـراف المعيــاري لتوزيــع معاينة $\bar{س}$.

ثانيا : ما هو شكل توزيع معاينة $\bar{س}$ ؟ ولماذا ؟

---

٨ ــ إذا كان حجم عبوات الأرز المعبأة في أحد المصانع يتوزع توزيعــا معتدلا وسطه الحسابي ٥ كجم وانحرافه المعيــاري ٠,٠٥ كجــم . وإذا سحبت عينة عشوائية من ٢٥ عبوة من إنتاج هـذا المصنــع ، فالمطلوب :

أولا : احتمال أن يكون الوسط الحسابي للعينة أقل من ٤,٨ كجم

ثانيا : احتمال أن يكون الوسط الحسابي للعينة أكبر من ٤,٩ كجم

ثالثا : احتمال أن ينحصر الوسط الحسابي للعينة بيــن ٤,٦ كجـم و٤.٧ كجم .

٩ ــ في أحد المجتمعات كان الوسط الحسابي لدخل الفرد الســنوي هــو ٦٠٠٠ جنيها ، بانحراف معياري ٢٠٠٠ جنيها ، وكان توزيع الدخل ملتويا جهة اليمين . فإذا سحبت عينة عشــوائية مــن ٤٠٠ فــرد ، فالمطلوب إيجاد احتمال أن يكون الوسط الحسابي لهذه العينة

أولا : أقل من ٤٥٠٠ جنيها .

ثانيا : بين ٤٥٠٠ جنيها و ٧٥٠٠ جنيها .

ثالثا : في حدود ١٠٠٠ جنيه من متوسط المجتمع .

رابعا : أقل من متوسط المجتمع بمبلغ ٥٠٠ جنيه أو أكثر .

١٠ ــ في أحد المدن كان توزيع فواتير الكهرباء توزيعا ملتويا ، وكان وسطه الحسابي ٧٠ جنيها بانحراف معياري ٣٠ جنيه . فإذا سحبت عينة عشوائية من ١٠٠ أسرة من هذه المدينة ، فالمطلوب إيجاد احتمال أن يكون الوسط الحسابي لهذه العينة

أولا : أكثر من ٧٥ جنيها .

ثانيا : بين٥٣ جنيها و٥٨ جنيها .

ثالثا : في حدود ٨ جنيهات من الوسط الحسابي للمجتمع .

رابعا : أكثر من متوسط المجتمع بمبلغ ١٠ جنيهات على الأقل .

١١ ــ ينتج أحد المصانع مصابيح كهربائية ، ومن المعروف أن الانحراف المعياري لعمر هذه المصابيح هو ١٢٠ ساعة . ولقد سحبت عينة عشوائية من ١٠٠ مفردة ووجد أن الوسط الحسابي لعمر المصابيح في هذه العينة هو ١٥٠٠ ساعة . والمطلوب : ما هو احتمال أن يكون عمر المصباح في هذه العينة في حدود ٢٠ ساعة من متوسط عمر المصابيح في هذا المصنع ؟

١٢ ــ في إحدى المحافظات كانت نسبة المدرسات (الإناث) في المدارس الحكومية ٦٠ ٪ . فإذا سحبت عينة عشوائية من ٨٠ من مدرسين هذه المدارس ، وكانت نسبة المدرسات (الإناث) في هذه العينة ق ، فالمطلوب :

أولا : حساب الوسط الحسابي والانحراف المعياري لتوزيع ق

ثانيا : ما هو شكل توزيع معاينة ق ؟

١٣ ـ في إحدى الامتحانات كانت نسبة الناجحين ٨٥ ٪ . وسحبت عينة عشوائية من ٥٠ طالب . فإذا كانت ق هي نسبة الناجحين في هذه العينة فالمطلوب :

أولا : حساب الوسط الحسابي والانحراف المعياري لتوزيع ق

ثانيا : ما هو شكل توزيع معاينة ق ؟

١٤ ـ سجلت إدارة المرور في إحدى المحافظات أن نسبة السيدات اللاتي تمتلكن رخص قيادة هي ٤٠ ٪ . ولقد سحبت عينة من ١٠٠ رخصة قيادة وكانت نسبة السيدات فيها هي ق . والمطلوب إيجاد احتمال أن تكون ق :

أولا : أقل من ٠,٣٨

ثانيا : تنحصر بين ٠,٣٧ و ٠,٤٣

ثالثا : في حدود ٠,٠٥ من نسبة المجتمع .

١٥ ـ يقوم أحد المصانع بإنتاج بطاريات للسيارات . وتدعي الإدارة أن ٩٠ ٪ من إنتاج المصنع مطابق للمواصفات . فإذا سحبت عينة عشوائية من ١٠٠ بطارية من إنتاج هذا المصنع وكانت نسبة البطاريات المطابقة للمواصفات هي ق ، فالمطلوب :

أولا : ما هو احتمال أن تكون نسبة العينة في حدود ٠,٠٥ من نسبة المجتمع ؟

ثانيا : ما هو احتمال أن تكون نسبة العينة أقل من نسبة المجتمع بمقدار ٠,٠٢ أو أكثر ؟

ثالثا : ما هو احتمال أن تكون نسبة العينة أكبر من نسبة المجتمع بمقدار ٠,٠٣ أو أكثر ؟

١٦ ــ من المعروف أن نسبة التالف من إنتاج أحد الآلات هو ٧ ٪ . ويقوم مفتش الرقابة على الجودة كل أسبوع بسحب عينة عشوائية من ١٠٠ منتج من إنتاج هذه الآلة . فإذا كانت نسبة التالف ٩ ٪ أو أكثر ، فإن المفتش يقرر إيقاف الإنتاج وتصليح الآلة . والمطلوب : ما هو احتمال أن يقرر المفتش إيقاف الإنتاج وتصليح الآلة بعد سحب عينة من ١٠٠ مفردة ؟

الفصل الثاني

تقدير معالم المجتمع

Estimation of the Population Parameters

مقدمة :

سبق وعرفنا في الفصل السابق المجتمع بأنه " جميع المفردات محل الدراسة سواء كانت في شكل إنسان أو حيوان أو جماد أو أشياء غير ملموسة " . كما عرفنا العينة بأنها " مجموعة من مفردات المجتمع " . وبينا أن أي مقياس إحصائي في المجتمع يسمى " معلمة " وأن أي مقياس إحصائي في العينة يسمى " إحصائية " .

ولقد سبق وبينا أن دراسة جميع مفردات المجتمع ــ والمسماة بالحصر الشامل ــ تتطلب تكاليف باهظة وتستهلك كثيراً من الوقت والجهد ، وقد يؤدي هذا الأسلوب إلى تلف الوحدات محل الدراسة . وبالإضافة إلى هذه الصعوبات ، فإن الحصول على معالم المجتمع من الأمور التي يتعذر الوصول إليها إن لم يكن ذلك مستحيلاً . لذلك لجأ الإحصائيون إلى استخدام المقاييس الإحصائية الناتجة من العينة ، أي " إحصائيات " العينة للتعرف على معالم المجتمع ، وهذا ما يسمى بالاستدلال الإحصائي Statistical Inference . وعندما نذكر العينة هنا نعني العينة العشوائية . ومن ثم يمكن تعريف الاستدلال الإحصائي بوصفه " مجموعة الأساليب الإحصائية التي بمقتضاها يمكننا الاستدلال على معالم المجتمع باستخدام إحصائيات عينة عشوائية من هذا المجتمع " .

وينقسم الاستدلال الإحصائي إلى موضوعين : تقدير معالم المجتمع واختبارات الفروض الإحصائية . ويتناول هذا الفصل دراسة تقدير معالم المجتمع بينما يختص الفصل الثالث بدراسة اختبارات الفروض الإحصائية . وهناك أسلوبان لتقدير معالم المجتمع ، التقدير بنقطة Point estimation والتقدير بفترة Interval estimation . وفي دراستنا لتقدير معالم المجتمع في الفصل الحالي سنتعرض للمباحث الآتية : التقدير بنقطة ، والتقدير بفترة ، ثم تقدير فترة ثقة لمتوسط مجتمع في حالة العينات الكبيرة ، وتقدير فترة ثقة لمتوسط مجتمع في حالة العينات الصغيرة ، فتقدير فترة ثقة لنسبة مجتمع ، فتقدير حجم العينة لتقدير متوسط مجتمع ، فتقدير حجم العينة لتقدير نسبة مجتمع .

(٢ ــ ١) أسلوب التقدير بنقطة :

وكما وضحنا في تعريف الاستدلال الإحصائي فإن عملية تقدير معالم المجتمع تتم عن طريق إحصائيات العينة ، فمثلاً إذا أردنا معرفة متوسط أعمار الطلبة في كلية التجارة يمكن أخذ عينة من ١٠٠ طالب وحساب الوسط الحسابي لعمرهم والانحراف المعياري لهذا العمر ، وليكن مثلاً الوسط الحسابي لأعمارهم $\bar{س}$ = ٢٠ سنة بانحراف معياري ع = ٣ سنوات ، فنقول هنا أن القيمة ٢٠ سنة هي تقدير نقطة للوسط الحسابي للمجتمع μ وأن القيمة ٣ سنوات هي تقدير نقطة للانحراف المعياري للمجتمع σ . ومن ثم فإن قيمة إحصائية العينة هي تقدير نقطة لمعلمة المجتمع . ومن ناحية أخرى فإن الإحصائية المستخدمة لتقدير معلمة المجتمع ــ والتي يمكن التعبير عنها بصيغة رياضية تبين الطريقة التي يتم بها حساب تقدير النقطة ــ تسمى "مقدر النقطة" Point estimator . فمثلاً الوسط الحسابي $\bar{س}$ هو مقدر نقطة للوسط الحسابي للمجتمع μ ، حيث : $\bar{س} = \frac{مجـ س}{ن}$.

وبالمثل فإن الانحراف المعياري في العينة ع هو مقدر نقطة للانحراف المعياري للمجتمع σ ، والنسبة في العينة ق هي مقدر نقطة لنسبة المجتمع θ . ولقد جرى العرف على استخدام علامة (^ هات) فوق رمز المعلمة للدلالة على مقدر النقطة فمثلاً إذا أردنا الدلالة على مقدر نقطة لمعلمة المجتمع م ، فإننا نكتب $\hat{م}$. بمعنى آخر فإن (^ هات) ترمز لمقدر المعلمة التي تكتب تحتها . وبالمثل فإن :

$\hat{\mu}$ هي مقدر نقطة للمعلمة μ .

$\hat{\sigma}$ هي مقدر نقطة للمعلمة σ .

$\hat{\theta}$ هي مقدر نقطة للمعلمة θ . وهكذا

ولتلخيص ما جاء في هذا المبحث نأخذ المثال السابق الخاص بأعمار طلاب كلية التجارة ، ومنه نجد أن :

اسم المعلمة	رمز المعلمة	مقدر نقطة	تقدير نقطة
الوسط الحسابي	μ	$\hat{\mu}$ = $\overline{س}$	$\hat{\mu}$ = ٢٠ سنة
الانحراف المعياري	σ	$\hat{\sigma}$ = ع	$\hat{\sigma}$ = ٣ سنوات

ومن الجدير بالذكر أن مقدر النقطة هو متغير عشوائي له توزيع احتمالي هو توزيع المعاينة الخاصة به ، بينما تقدير النقطة هو مقدار ثابت .

وعند اختيار مقدر نقطة ما ، يثار التساؤل عما إذا كان هذا المقدر " جيد " أم لا . فعلى سبيل المثال : هل الوسط الحسابي مقدر جيد للمعلمة μ ؟ أم هل من الأفضل استخدام الوسيط ؟ للإجابة على هذه التساؤلات يجب الأخذ في الاعتبار بعض المعايير وهي : عدم التحيز ، والكفاءة ، والكفاية ، والاتساق ، ومتوسط مربع الخطأ .

١ ـ عدم التحيز : Unbiasedenss

يقال أن مقدر نقطة هو مقدر غير متحيز لمعلمة المجتمع إذا كان متوسط الإحصائية ـ المحسوبة من جميع العينات العشوائية الممكنة المحسوبة من هذا المجتمع والتي لها نفس الحجم ـ يساوي معلمة المجتمع . أو بعبارة أخرى فإن مقدر نقطة يقال عنه أنه غير متحيز إذا كان متوسط توزيعه العيني يساوي معلمة المجتمع . ومن ثم فإن $\hat{م}$ هو مقدر نقطة غير متحيز للمعلمة م إذا كان :

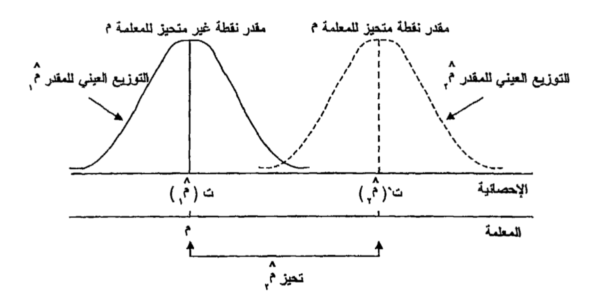
$$ت (\hat{م}) = م \qquad (١)$$

وإذا كانت $\hat{م}$ مقدر نقطة متحيز ، فإن مقدار التحيز (أ) يقاس كما يلي :

$$أ = ت (\hat{م}) - م \qquad (٢)$$

ويبين شكل (١) مقدري نقطة : $\hat{م}_١$ مقدر نقطة غير متحيز ، $\hat{م}_٢$ مقدر نقطة متحيز . ويوضح الرسم أن $\hat{م}_٢$ تعطي تقديرات بعيدة عن المعلمة م ، في حين أن $\hat{م}_١$ تعطي تقديرات قريبة من المعلمة م . ويعتبر المقدر $\hat{م}_٢$ متحيز لأن ت ($\hat{م}_٢$) ≠ م . ويقاس التحيز بالمقدار أ وهو الفرق بين ت ($\hat{م}_٢$) و م أي أن :

$$أ = ت (\hat{م}_٢) - م$$



شكل (١) : توزيعي معاينة مقدر غير متحيز وآخر متحيز

وقد يكون المقدر المتحيز مقدراً مرغوباً فيه إذا كان مقدار التحيز صغيرا ، طالما أن هذا المقدر يتمتع بخصائص أخرى مرغوب فيها .

ولقد تبين من دراستنا في المعاينة الإحصائية أن الوسط الحسابي لتوزيع المعاينة للمتغير العشوائي يساوي الوسط الحسابي للمجتمع μ . ويمكن استخدام الرمز ت ($\overline{س}$) ـ أي توقع $\overline{س}$ ـ للدلالة على الوسط الحسابي لتوزيع المعاينة للمتغير العشوائي $\overline{س}$. أي أن :

$$ت\ (\overline{س}) = \mu$$

إذن الوسط الحسابي للعينة ($\overline{س}$) يعتبر مقدر غير متحيز للوسط الحسابي للمجتمع μ .

أما بالنسبة للوسيط ، فإذا كان التوزيع ملتوياً فإن الوسيط المحسوب من العينة يعتبر تقديراً متحيزاً للوسط الحسابي μ ، أي أن : ت (الوسيط) $\neq \mu$. فعلى سبيل المثال الوسيط المحسوب من عينة عشوائية لدخل مجموعة من

الأسر يكون أقل بكثير من الوسط الحسابي لدخل الأسرة في المجتمع ، وهـــذا لأنه عادة يكون توزيع الدخل ملتوياً ناحية اليمين .

أما بالنسبة للتباين المأخوذ من العينة والذي يحسب من المعادلة :

$$ع^{٢} = \frac{١}{ن - ١} \sum (س - \bar{س})^{٢}$$

فهو يعتبر مقدر نقطة غير متحيز للمعلمة $\sigma^{٢}$ لأن :

$$ت(ع^{٢}) = \sigma^{٢}$$

ويلاحظ أننا قسمنا هنا على ن – ١ (وهي درجات الحرية)[١] بـــدلاً من القسمة على ن .

أما الانحراف المعياري للعينة فهو مقدر نقطة متحيز للمعلمة σ ، وذلك لأن :

$$ت(ع) \neq \sigma$$

٢ ـــ الكفاءة : Efficiency

تقاس كفاءة مقدر غير متحيز عن طريق تباين توزيعه العيني . فـــإذا كان لدينا مقدري نقطة غير متحيزين من عينتين لهما نفس الحجم ، فإن مقـــدر النقطة الأعلى كفاءة نسبياً هو ذلك المقدر ذو التباين الأصغر .

---

[١] يمكن تعريف درجات الحرية بأنها عدد المشاهدات التي يمكن لختيارها بحرية ، أو عـــدد المتغـيرات التـي يمكن أن تتغير بحرية ، أ عدد المتغيرات المستقلة . ففي حالة وجود مجموع مربعات كميات معينـــة ، فإن درجات الحرية تعرف بأنها عدد المربعات ناقص عدد المتغيرات المستقلة المفروضة على الكميـات محل البحث . . فعند ن من المشاهدات لدينا ن من مربعات انحرافات القيم عـــن وســطها الحســابي ، ولكن هناك ن – ١ فقط منها هي المستقلة ، بمعنى أنه إذا عرفنا ن – ١ من هذه الانحرافات نســتطيع تحديد الانحراف النوني . والسبب يرجع إلى أن لدينا قيداً وهو : أن مجموع الانحرافات عـــن الوسـط الحسابي يجب أن يساوي صفر . ومن ثم فإن مجموع مربعات انحرافات ن من القيـــم عـــن وســطها الحسابي لها ن – ١ درجات حرية .

ومن ثم ، فإذا كان لدينا مقدري نقطة غير متحيزين $\hat{م}_{١}$ ، $\hat{م}_{٢}$ للمعلمة م ، فإن $\hat{م}_{١}$ يكون أكثر كفاءة نسبياً إذا كان :

$$\sigma^{٢}(\hat{م}_{١}) < \sigma^{٢}(\hat{م}_{٢}) \quad \text{حيث} \quad ت(\hat{م}_{١}) = ت(\hat{م}_{٢}) = م \qquad (٣)$$

مثال (١) :

المطلوب مقارنة كلا من الوسط الحسابي والوسيط من حيث الكفاءة كمقدري نقطة غير متحيزين للوسط الحسابي لمجتمع يتوزع توزيعاً معتدلاً ، علماً بأن كلاً من الوسط الحسابي والوسيط غير متحيزين وأن :

$$\text{تباين } \bar{س} = \frac{\sigma^{٢}}{ن}$$

$$\text{تباين الوسيط} = ١,٥٧ \frac{\sigma^{٢}}{ن}$$

تباين الوسيط أكبر من تباين الوسط الحسابي ، مع افتراض نفس حجم العينة . ومن ثم فإن الوسط الحسابي أكثر كفاءة من الوسيط . ويمكن قياس الكفاءة النسبية لمقدر ما بالنسبة لمقدر آخر عن طريق النسبة بين تبايني المقدرين . فإذا كان المقدرين غير متحيزين فإن :

$$\text{الكفاءة النسبية} = \frac{\text{تباين } \hat{م}_{١}}{\text{تباين } \hat{م}_{٢}} \qquad (٤)$$

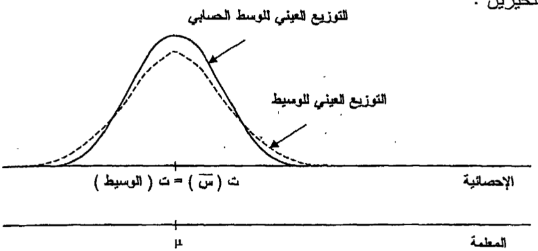
وفي مثالنا السابق :

$$\text{الكفاءة النسبية} = \frac{\text{تباين } \bar{س}}{\text{تباين الوسيط}}$$

$$= \frac{\frac{\sigma^{٢}}{ن}}{١,٥٧ \frac{\sigma^{٢}}{ن}} = ٠,٦٤$$

وبما أن هذه النسبة أقل من الواحد الصحيح ، إذن تباين $\bar{س}$ أقل من تباين الوسيط ومن ثم فإن لوسط الحسابي أكثر كفاءة من الوسيط .

ويوضح شكل (٢) هذه الحالة . ومن الملاحظ أن المقدرين غير متحيزين .



شكل (٢) : التوزيع العيني للوسط الحسابي مقارن بالتوزيع العيني للوسيط من حيث الكفاءة بوصفهما مقدري نقطة غير متحيزين للمعلمة μ

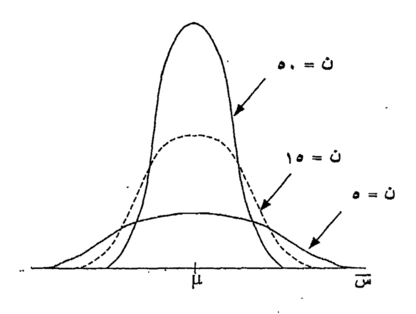
ولكن التوزيع العيني للوسط الحسابي أكثر تركزا حول μ عن التوزيع العيني للوسيط

٣ ــ الاتساق : Consistency

يقال عن إحصائية أنها مقدرا متسقا إذا اقترب المقدر من معلمة المجتمع مع ازدياد حجم العينة . ولقد رأينا أن $\sigma^{2}_{\bar{س}} = \frac{\sigma^{2}}{ن}$. وهذا يعني أنه بازدياد حجم العينة ن فإن تباين توزيع معاينة $\bar{س}$ ($\sigma^{2}_{\bar{س}}$) سينخفض ، أي أن الأوساط الحسابية لعينات ($\bar{س}_{1}$ ، $\bar{س}_{2}$ ، ... ، $\bar{س}_{ن}$) تقترب من الوسط الحسابي للمجتمع μ مع ازدياد حجم العينة . ومن ثم فإن الوسط الحسابي $\bar{س}$ هو مقدر متسق للمعلمة μ .

ويكون المقدر متسقاً إذا كان تباينه يقترب من الصفر عندما يقترب حجم العينة من ∞ . وهنا أيضاً نجد أن الوسط الحسابي $\bar{س}$ يحقق هــذا الشــرط لأن $\frac{\sigma^2}{ن}$ تقترب من الصفر عندما ن تقترب من ∞ . ويمكن أيضاً تبيـــان أن $ع^2$ هو مقدر متسق للمعلمة σ^2 .

ويبين شكل (٣) أن $\bar{س}$ هو مقدر متسق للمعلمـــة μ عندمـــا يـــكون التوزيع العيني معتدل ، ويبين الشكل أن $\bar{س}$ تزداد اقتراباً من μ كلما ازداد حجم العينة .



شكل (٣) : $\bar{س}$ مقدر متسق للمعلمة μ عندما يكون التوزيع العيني معتدل

٤ ـــ الكفاية : Sufficiency

يقال عن مقدر أنه كاف إذا استخدم جميع البيانات الموجودة في العينـــة والخاصة بحساب المعلمة المراد تقديرها . وإذا كان هناك مقدر كاف فلا يكـــون مجدياً استخدام أي مقدر آخر أقل كفاية ، فالمقدر الكاف يستخدم كل المعلومـــات الموجودة في العينة والخاصة بتقدير المعلمة . ويعتبر الوسط الحسابي $\bar{س}$ مقدر

كاف للمعلمة μ لأنه يستخدم في حسابه نفس المعلومات التي تستخدمها المعلمة μ . فلحساب μ نجمع جميع القيم ونقسم على عددها . ونفعل نفس الشيء في العينة لحساب $\bar{س}$. أما الوسيط فهو لا يعتبر مقدر كاف للمعلمة μ ، فلحساب الوسيط نوجد ترتيب الوسيط ، ثم نوجد القيمة الوسطى ، ولا نستخدم جميع القيم كما يحدث عند حساب μ . هذا وتعتبر النسبة ق مقدر كاف للمعلمة θ لأنها تستخدم في حسابها نفس المعلومات التي تستخدمها المعلمة θ . فلحساب النسبة θ في المجتمع نقسم عدد المفردات التي لها نفس الصفة على عدد المفردات الكلية . ونفعل نفس الشيء في العينة لحساب النسبة ق .

٥ ــ متوسط مربع الخطأ : Mean squared error

يمزج هذا المعيار معيار عدم التحيز ومعيار الكفاءة ، وهذا المعيار مفيد في حالة مقارنة مقدرين أحدهما أو كلاهما متحيز . ويمزج متوسط مربع الخطأ للمقدر $\hat{م}$ تباين التوزيع العيني للمقدر $\hat{م}$ أي $\sigma^{٢}$ ($\hat{م}$) وتحيز المقدر أي ت ($\hat{م}$) ـ $\hat{م}$.

ويمكن تعريف متوسط مربع الخطأ للمقدر م كالآتي :

$$\text{متوسط مربع الخطأ} = \sigma^{٢}(\hat{م}) + [\,ت\,(\hat{م}) - \hat{م}\,]^{٢} \qquad (٥)$$

وعند مقارنة مقدرين ، فإن المقدر الذي له أقل متوسط مربع الخطأ يقال عنه أنه ذو كفاءة نسبية أعلى في متوسط مربع الخطأ عن المقدر الآخر .

ووفقا لهذا المعيار يفضل المقدر المتحيز الذي يتمتع بتوزيع عيني مركز حول المعلمة م ، على مقدر غير متحيز له توزيع عيني ذو تشتت أكبر .

مثال (٢) : فإذا كان لدينا المقدرين الآتين :

المقدر	التباين	التحيز	متوسط مربع الخطأ
$\hat{م}$	$\sigma^{٢}(\hat{م})$	$ت(\hat{م}) - م$	$\sigma^{٢}(\hat{م}) + [ت(\hat{م}) - م]^{٢}$
$\hat{م}_{١}$	١٠	٣	$١٠ + ٣^{٢} = ١٩$
$\hat{م}_{٢}$	٤٠	٠	٤٠ + ٠ = ٤٠

وطبقا لمعيار متوسط مربع الخطأ يفضل المقدر $\hat{م}_{١}$ على $\hat{م}_{٢}$.

وبالإضافة إلى المعايير السابقة فإنه هناك معيارين آخرين لمعرفة إذا كان المقدر جيدا أم لا ، وهما : طريقة الإمكان الأكبر Maximum Likelihood Method وطريقة المربعات الصغرى Method of Least Squares . ولن نتعرض لهما في دراستنا .

(٢ ــ ٢) أسلوب التقدير بفترة :

من دراستنا للتقدير بنقطة تبين لنا أن الوسط الحسابي $\bar{س}$ المحسوب من العينة هو مقدر جيد لمعلمة المجتمع μ . ولكن هذا لا يعني أن $\bar{س}$ تساوي μ بالضبط ، فقد تأخذ $\bar{س}$ قيم أقل من أو أكبر من μ طبقا للعينة المحسوبة منها . ومن ثم فإن التقدير بنقطة يعطي قيمة لمعلمة المجتمع تكون في أغلب الأحيان مختلفة تماما عن القيمة الحقيقية لهذه المعلمة . لذلك يكون من الأفضل استخدام أسلوب التقدير بفترة . وطبقا لهذا الأسلوب فإنه يتم وضع فترة حول مقدر النقطة بحيث يكون من المحتمل أن تحتوي هذه الفترة على معلمة المجتمع باحتمال محدد مقدما وهو ما يسمى بدرجة الثقة degree of confidence أو مستوى الثقة confidence level . ويمكننا الحصول على هذه الفترة لتحقيق

أي درجة ثقة مطلوبة لذلك سميت هذه الفترة بفترة الثقة confidence interval كما سمي حدي هذه الفترة بحدي الثقة confidence limits . ولإيضاح ذلك سنأخذ الحالة القصوى ، فيمكننا القول بدرجة ثقة ١٠٠ ٪ بأن الفترة من $-\infty$ إلى $+\infty$ تشتمل على معلمة المجتمع ، وهذه الفترة ليس لها في الواقع أهمية عملية . ويمكننا تضييق هذه الفترة ولكن يتم هذا بثمن ألا وهو تخفيض درجة الثقة بأن هذه الفترة تحتوي على المعلمة المجهولة .

وجملة القول : كل فترة مصحوبة بدرجة ثقة معينة لذلك سميت بفترة الثقة . وتحدد درجة الثقة ــ المصاحبة لفترة الثقة ــ مدى الثقة التي تكون لدينا بأن هذه الفترة تحتوي على معلمة المجتمع الحقيقية . ويرمز لدرجة الثقة بالرمز ١٠٠ (١ − α) ٪ ، حيث α هي الحرف الإغريقي ويقرأ " ألفا " ، وتسمى α بمستوى المعنوية significance level ، كما يسمى الاحتمال (١ − α) بمعامل الثقة confidence coefficient وعادة يتم استخدام درجات الثقة ٩٠ ٪ أو ٩٥ ٪ أو ٩٩ ٪ ، فعلى سبيل المثال إذا قلنا أن :

ح (٣٠ $\leq \mu \leq$ ٥٠) = ٠,٩٩

فإن هذا يعني أنه باحتمال قدره ٩٩ ٪ **تحتوي الفترة** ما بين ٣٠ ، ٥٠ **على القيمة الحقيقية لمتوسط المجتمع** μ . ولكن من الخطأ القول بأنه باحتمال ٩٩ ٪ تنحصر μ بين ٣٠ و ٥٠ وذلك لأن معلمة المجتمع قيمة ثابتة والاحتمالات تتعلق دائماً بالمتغيرات العشوائية ، فالاحتمال هنا هو معامل الثقة (٩٩ ٪) بينما المتغيرات العشوائية هي حدي الثقة (٣٠ ، ٥٠) .

هذا ويفضل بعض الإحصائيون مناقشة تقدير معالم المجتمع والفروض الإحصائية بمعلومية أو عدم معلومية σ . ولكن في دراستنا هنا يتم استخدام معيار العينات الكبيرة والعينات الصغيرة ، والسبب في ذلك يرجع إلى أن الانحراف المعياري للمجتمع يكون في غالب الأحيان دائماً مجهولاً . لذلك

فدراسة تقدير معالم المجتمع واختبارات الفروض ، طبقا لما إذا كانت العينات كبيرة أم صغيرة ، تكون أكثر واقعية من كون σ معلومة أو غير معلومة .

(٢ ـ ٣) تقدير فترة ثقة لمتوسط مجتمع في حالة العينات الكبيرة :

سبق وذكرنا عند دراستنا للمعاينة الإحصائية أنه إذا سحبنا عينة عشوائية حجمها ن من مجتمع معين وحسبنا الوسط الحسابي منها $\bar{س}_١$ ، ثم سحبنا عينة عشوائية أخرى من نفس المجتمع حجمها ن وحسبنا منها الوسط الحسابي $\bar{س}_٢$ وهكذا . . . إلى أن يتم سحب جميع العينات العشوائية الممكنة التي لها نفس الحجم ن من هذا المجتمع ، نجد أن $\bar{س}$ متغير عشوائي له توزيع احتمالي هو التوزيع العيني للوسط الحسابي $\bar{س}$. وطبقا لنظرية النهاية المركزية ، فعندما تكون العينة كبيرة فإن التوزيع العيني للوسط الحسابي $\bar{س}$ يتبع تقريبا التوزيع المعتدل الذي وسطه الحسابي μ وتباينه $\frac{\sigma^٢}{ن}$ ، وهذا بصرف النظر عن شكل التوزيع الأصلي للمجتمع . وتعتبر العينة كبيرة إذا كان حجمها ن $\geq$ ٣٠ .

ويجب التفرقة هنا بين حالة ما إذا كانت σ معلومة أم غير معلومة .

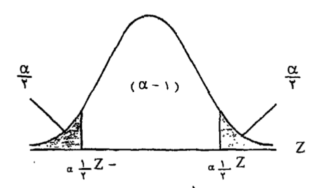
أ ـ إذا كانت σ معلومة :

وبما أن التوزيع العيني يتبع التوزيع المعتدل ع (μ ، $\frac{\sigma^٢}{ن}$) ، ومن ثم فإن المتغير العشوائي Z يتبع التوزيع المعتدل المعياري بوسط حسابي صفر وانحراف معياري ١ ، أي ع (٠ ، ١) ، حيث :

$$Z = \frac{\bar{س} - \mu}{\frac{\sigma}{\sqrt{ن}}}$$

حيث $\frac{\sigma}{\sqrt{ن}}$ هو الخطأ المعياري .

فإذا تم اختيار درجة ثقة ١٠٠ (١ − α) ٪ ، فهذا يعني أن معامل الثقة (أي الاحتمال) هو (١ − α) . ومن ثم فإن المساحة الباقية تحت المنحنى [١ − (١ − α) = α] تقسم بالتساوي بين طرفي التوزيع أي ($\frac{\alpha}{2}$) . وكما هو موضح في شكل (٤) فإن (١ − α) من المساحة الكلية تقع بين $-Z_{\frac{1}{2}\alpha}$ ، $Z_{\frac{1}{2}\alpha}$. حيث $Z_{\frac{1}{2}\alpha}$ هي قيمة Z التي تكون المساحة إلى يمينها مساوية $\frac{1}{2}\alpha$. وتستخرج هذه القيمة من جدول التوزيع المعتدل المعياري ، وبما أن هذه الجداول تعطي فقط قيم Z الموجبة ، فإننا سنرمز إلى قيمة Z التي تكون المساحة على يسارها مساوية $\frac{1}{2}\alpha$ بالرمز $-Z_{\frac{1}{2}\alpha}$.
وبالتالي فإن :



شكل (٤) : التوزيع المعتدل المعياري مبينا حدي الثقة عليه

$$ح\left(-Z_{\frac{1}{2}\alpha} \le Z \le Z_{\frac{1}{2}\alpha}\right) = 1 - \alpha$$

$$ح\left(-Z_{\frac{1}{2}\alpha} \le \frac{\bar{س} - \mu}{\frac{\sigma}{\sqrt{ن}}} \le Z_{\frac{1}{2}\alpha}\right) = 1 - \alpha \qquad (٦)$$

بضرب طرفي المتباينة في $\frac{\sigma}{\sqrt{ن}}$ (حيث $\frac{\sigma}{\sqrt{ن}}$ مقدار موجب) ،

$$ح\left(-Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}} \le \bar{س} - \mu \le Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}}\right) = 1 - \alpha$$

وبطرح $\bar{س}$ من طرفي المتباينة :

$$ح\left(-\bar{س} - Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}} \leq -\mu \leq -\bar{س} + Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}}\right) = 1 - \alpha$$

وبضرب طرفي المتباينة في –١ :

$$ح\left(\bar{س} + Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}} \geq \mu \geq \bar{س} - Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}}\right) = 1 - \alpha \quad (٧)$$

وهذا يعني أن هناك احتمالا قدره ($1 - \alpha$) بأن الفترة التي يكون

حدها الأدنى = $\bar{س} - Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}}$ ، حدها الأعلى = $\bar{س} + Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}}$

تحتوي على قيمة المعلمة μ

ويسمى المقدار ($Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}}$) بخطأ التقدير Error of estimate [1] .

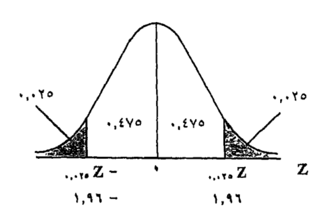
وجملة القول :

> $$\bar{س} \pm Z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{ن}} \quad (٨)$$
>
> هو مقدر فترة ثقة ١٠٠ ($1 - \alpha$) ٪ لمتوسط مجتمع معلوم التباين
>
> عندما ن $\geq$ ٣٠

لإيضاح كيفية الحصول على قيمة $Z_{\frac{1}{2}\alpha}$ من جدول التوزيع المعتدل المعياري ، فعلى سبيل المثال إذا كان المطلوب تحديد فترة ثقة ٩٥ ٪ للمعلمة μ ، فهذا يعني أن المساحة تحت المنحنى المعتدل بين نقطتين هي ٩٥ ٪ أي ٠,٠٥ على يمين ويسار μ معا ، ومن ثم ٠,٠٢٥ على كل من يمين ويسار μ ، كما هو مبين في شكل (٥) .

---

[1] يجب ملاحظة أن كلمة خطأ هنا لا تعني المعنى المعروف لها . ولكن يقصد بها الاختلافات في قيم الإحصائيات من عينة إلى أخرى .



شكل (٥) : إيجاد قيمة $Z_{0.025}$

إذا أخذنا جدول المنحنى المعتدل المعياري الذي يعطي المساحة من $-\infty$ إلى Z ، نتبع الخطوات التالية :

١ ــ نوجد قيمة α ، بما أن ١ - α = ٠,٩٥ ، إذن α = ٠,٠٥

٢ ــ نوجد قيمة $\frac{1}{2}\alpha$ ، أي $\frac{0.05}{2}$ = ٠,٠٢٥

٣ ــ نوجد (١ - $\frac{1}{2}\alpha$) ، أي : ١ - ٠,٠٢٥ = ٠,٩٧٥

٤ ــ نبحث في الجدول نفسه أي في المساحة حتى نجد قيمة ٠,٩٧٥ ، فنجد أن هذه القيمة تقع أمام Z = ١,٩ وتحت Z = ٠,٠٦ وبالتالي قيمة Z تساوي ١,٩٦ .

ويبين الجدول التالي قيم $Z_{\frac{1}{2}\alpha}$ لدرجات ثقة مختلفة شائعة الاستخدام :

درجة الثقة ١٠٠ (١ - α) ٪	١ - α	α	$\frac{1}{2}\alpha$	$Z_{\frac{1}{2}\alpha}$
٪ ٩٠	٠,٩٠	٠,١٠	٠,٠٥	١,٦٤٥
٪ ٩٥	٠,٩٥	٠,٠٥	٠,٠٢٥	١,٩٦٠
٪ ٩٩	٠,٩٩	٠,٠١	٠,٠٠٥	٢,٥٧٥

مثال (٣) :

سحبت عينة عشوائية من ١٠٠ مصباح كـــهربائي مـــن إنتـــاج أحـــد المصانع فوجد أن الوسط الحسابي لعمر المصباح ١٠٠٠ ساعة . فإذا علمت أن الانحراف المعياري لعمر المصباح في المجتمع هو ١٥٠ ساعة ، فالمطلوب :

(أ) إيجاد تقدير نقطة لمتوسط عمر المصباح .

(ب) إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط عمر المصباح في هذا المصنع .

(حـــ) إيجاد تقدير فترة ثقة ٩٩ ٪ لمتوسط عمر المصباح في هذا المصنع .

(د) إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط عمر المصباح إذا كان حجم العينـــة ٢٠٠ مصباح .

الحـــــل :

ن = ١٠٠ ، $\overline{س}$ = ١٠٠٠ ساعة ، σ = ١٥٠ ساعة ، ١ − α = ٠,٩٥ ،

α = ٠,٠٥ ، $\frac{١}{٢}\alpha$ = ٠,٠٢٥ ، $Z_{٠,٠٢٥}$ = ١,٩٦

(أ) تقدير نقطة لمتوسط عمر المصباح : $\hat{\mu} = \overline{س}$ = ١٠٠٠ ساعة .

(ب) بما أن ن > ٣٠ فيمكننا تطبيق نظرية النهاية المركزية ، ومن ثم فـــإن توزيع المعاينة يتبع التوزيع المعتدل بصرف النظر عن التوزيع الأصلـــي للمجتمع . ومن ثم فإن مقدر فـــترة ثقـــة ١٠٠ (١ − α) ٪ لمتوســـط مجتمع معلوم التباين هو :

$$\overline{س} \pm Z_{\frac{١}{٢}\alpha} \frac{\sigma}{\sqrt{ن}}$$

$$١٠٠٠ \pm ١,٩٦ \frac{١٥٠}{\sqrt{١٠٠}}$$

$$١٠٠٠ \pm ٢٩,٤$$

الحد الأدنى لفترة الثقة = ٩٧٠,٦ ساعة .

الحد الأعلى لفترة الثقة = ١٠٢٩,٤ ساعة .

وهذا يعني أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٩٧٠,٦ ساعة إلى ١٠٢٩,٤ ساعة على الوسط الحسابي لعمر المصباح في المجتمع .

(حـ) $1 - \alpha = ٠,٩٩$ ، $\alpha = ٠,٠١$ ، $\frac{١}{٢}\alpha = ٠,٠٠٥$ ، $Z_{٠,٠٠٥} = ٢,٥٧٥$

ويكون تقدير فترة الثقة ٩٩ ٪ لمتوسط عمر المصباح هو :

$$١٠٠٠ \pm ٢,٥٧٥ \frac{١٥٠}{\sqrt{١٠٠}}$$

$$١٠٠٠ \pm ٣٨,٦٢٥$$

الحد الأدنى لفترة الثقة = ٩٦١,٣٧٥ ساعة .

الحد الأعلى لفترة الثقة = ١٠٣٨,٦٢٥ساعة .

وهذا يعني أنه بدرجة ثقة ٩٩ ٪ تحتوي الفترة من ٩٦١,٤ ساعة إلى ١٠٣٨,٦ ساعة على الوسط الحسابي لعمر المصباح في المجتمع .

(د) ن = ٢٠٠

ويكون تقدير فترة الثقة ٩٥ ٪ لمتوسط عمر المصباح هو :

$$١٠٠٠ \pm ١,٩٦ \frac{١٥٠}{\sqrt{٢٠٠}}$$

$$١٠٠٠ \pm ٢٠,٧٨٩$$

الحد الأدنى لفترة الثقة = ٩٧٩,٢١١ ساعة .

الحد الأعلى لفترة الثقة = ١٠٢٠,٧٨٩ساعة .

وهذا يعني أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٩٧٩,٢ ساعة إلى ١٠٢٠,٨ ساعة على الوسط الحسابي لعمر المضباح في المجتمع .

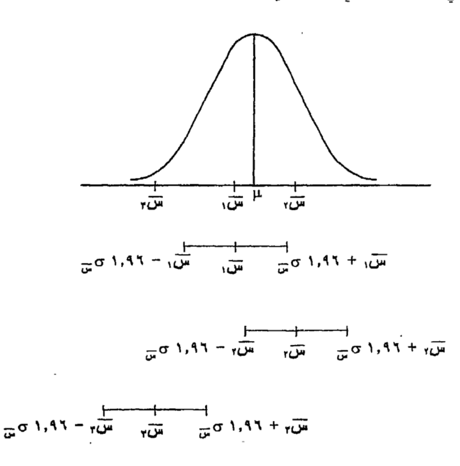
العوامل المحددة لفترة الثقة :

بالنظر إلى قانون (٨) نجد أن العوامل المحددة لفترة الثقة هي : $Z_{\frac{\alpha}{2}}$ ، ن ، σ . إلا أن الباحث لا يستطيع التحكم في الانحراف المعياري للمجتمع لكنه يستطيع التحكم في مستوى المعنوية وفي حجم العينة . ولقد رأينا في مثال (٣) : أنه عند درجة ثقة ٩٩ ٪ فإن فترة الثقة كانت (٩٦١,٤ ، ١٠٣٨,٦) وعند درجة ثقة ٩٥ ٪ فإن فترة الثقة كانت (٩٧٠,٦ ، ١٠٢٩,٤) وهذا يعني أنه إذا انخفضت درجة الثقة فإن فترة الثقة تصبح أُضيق ، كما رأينا أيضاً أنه إذا ازداد حجم العينة من ١٠٠ مفردة إلى ٢٠٠ مفردة ــ مع الاحتفاظ بنفس درجة الثقة ٩٥ ٪ فإن فترة الثقة تصبح أُضيق : من (٩٧٠,٦ ، ١٠٢٩,٤) إلى (٩٧٩,٢ ، ١٠٢٠,٨) . وعموماً يفضل الباحث فترة ثقة ضيقة لأن هذا يعطي نتيجة أكثر دقة بالنسبة لموقع معلمة المجتمع ، ويستطيع تحقيق ذلك إما بتخفيض درجة الثقة أو بزيادة حجم العينة . ولكن تخفيض درجة الثقة يخفض احتمال أن تحتوي فترة الثقة على معلمة المجتمع ، وهذا أمر غير مرغوب فيه . لذلك من الأفضل زيادة حجم العينة إذا ما أردنا تضييق فترة الثقة وبالتالي الحصول على نتائج أكثر دقة فيما يتعلق بموقع معلمة المجتمع .

تفسير درجة الثقة :

ما هو التفسير لدرجة ثقة ٩٥ ٪ مثلاً ؟ ففي المثال السابق إذا أخذنا جميع العينات العشوائية الممكنة التي حجمها ن من هذا المجتمع ، وحسبنا في كل منها الأوساط الحسابية ، ثم وضعنا فترة ثقة ٩٥ ٪ للمعلمة μ حول كل وسط حسابي ، يمكننا التوقع بأن ٩٥ ٪ من هذه الفترات ستحتوي على المعلمة μ و ٥ ٪ منها لا تحتوي عليها . ويبين شكل (٥) الأوساط الحسابية $\overline{س}_١$ و $\overline{س}_٢$ و $\overline{س}_٣$ لثلاث عينات عشوائية مسحوبة من نفس المجتمع ، كما يبين

الشكل فترات الثقة حول هذه المتوسطات . ومن الواضح في هـــذا الشـــكل أن فترات الثقة حول $\bar{س}_{١}$ ، $\bar{س}_{٢}$ تحتوي بداخلها على μ . ويمكـــن القـــول بأنـــه بدرجة ثقة ٩٥ ٪ إذا أخذنا جميع العينات الممكنة التي لها نفس الحجم من هـــذا المجتمع ووضعنا فترة ثقة ٩٥ ٪ حول الأوســـاط الحســـابية لـــهذه العينـــات ، فسيكون هناك ٩٥ ٪ من هذه الفترات مثل تلك الفترات التي حـــول $\bar{س}_{١}$ ، $\bar{س}_{٢}$ والتي تحتوي بداخلها على μ ، كما سيكون هناك ٥ ٪ من هذه الفترات مثل تلك الفترة التي حول $\bar{س}_{٣}$ فهي لا تحتوي بداخلها على μ .



شكل (٥) : التوزيع العيني للأوساط الحسابية $\bar{س}$
موضحا ثلاث فترات ثقة للمعلمة μ

ب ــ إذا كانت σ غير معلومة :

لقد افترضنا في دراستنا لفترات الثقة أن σ معلومة ، ولكن في كثير من الأحيان تكون σ غير معلومة ، ففي هذه الحالة نستخدم الانحراف المعياري للعينة ع كتقدير نقطة للمعلمة σ . وبما أن حجم العينة كبير (ن ≥ ٣٠) ، فطبقاً لنظرية النهاية المركزية فإن التوزيع العيني للأوساط الحسابية يخضع تقريباً للتوزيع المعتدل وبالتالي فإن مقدر فترة الثقة في هذه الحالة يصبح كما يلي :

$$\bar{س} \pm Z_{\frac{1}{2}\alpha} \frac{ع}{\sqrt{ن}} \qquad (٩)$$

هو مقدر فترة ثقة ١٠٠ (١ − α) ٪ لمتوسط مجتمع مجهول التباين عندما ن ≥ ٣٠ .

مثال (٤) :

سحبت عينة عشوائية من أجور ٥٠ عامل من عمال أحد المصانع ، وفيما يلي التوزيع التكراري لأجور هؤلاء العمال :

فئات الأجر (بالجنيهات)	١٠٠ −	١٥٠ −	٢٠٠ −	٢٥٠ −	٣٠٠ −	٣٥٠ وأقل من ٤٠٠	المجموع
عدد العمال	٥	١٠	١٥	٨	٧	٥	٥٠

والمطلوب :

أ ــ إيجاد تقدير نقطة لمتوسط أجر العمال في هذا المصنع .

ب ــ إيجاد تقدير فترة ثقة ٩٩ ٪ لمتوسط أجر العمال في هذا المصنع .

الحـــل :

فئات الأجر (بالجنيهات)	عدد العمال ك	مراكز الفئات س	س ك	س^2 ك
١٠٠ –	٥	١٢٥	٦٢٥	٧٨١٢٥
١٥٠ –	١٠	١٧٥	١٧٥٠	٣٠٦٢٥٠
٢٠٠ –	١٥	٢٢٥	٣٣٧٥	٧٥٩٣٧٥
٢٥٠ –	٨	٢٧٥	٢٢٠٠	٦٠٥٠٠٠
٣٠٠ –	٧	٣٢٥	٢٢٧٥	٧٣٩٣٧٥
٣٥٠ وأقل من ٤٠٠	٥	٣٧٥	١٨٧٥	٧٠٣١٢٥
المجموع	٥٠		١٢١٠٠	٣١٩١٢٥٠

١ ــ مقدر نقطة لمتوسط أجر العمال في هذا المصنع = $\hat{\mu} = \overline{\text{س}}$

$$\overline{\text{س}} = \frac{\text{مجـ س ك}}{\text{مجـ ك}} = \frac{١٢١٠٠}{٥٠} = ٢٤٢ \text{ جنيه}$$

٢ ــ بما أن حجم العينة كبير (ن = مجـ ك = ٥٠) ، فطبقا لنظريـــة النهايـة المركزية فإن التوزيع العيني للأوساط الحسابية يخضع للتوزيع المعتدل ، وحيث أن σ مجهولة فإننا نستخدم الانحراف المعياري للعينة ع كمقـــدر نقطة لمعلمة المجتمع σ . وحيث :[1]

$$\text{ع}^2 = \frac{١}{\text{مجـ ك} - ١}\left[\text{مجـ س}^2\text{ ك} - \frac{(\text{مجـ س ك})^2}{\text{مجـ ك}}\right]$$

$$= \frac{١}{٤٩}\left[٣١٩١٢٥٠ - \frac{(١٢١٠٠)^2}{٥٠}\right]$$

$$= ٥٣٦٨.٣٦٧$$

[1] لقد قمنا بالقسمة على ن – ١ (درجات الحرية) لأن الانحراف المعياري محسوب من عينة . إلا أنه يمكن القسمة على ن حيث أن العينة كبيرة .

$$ع = \sqrt{٥٣٦٨,٣٦٧}$$

= ٧٣,٢٦٩ جنيهاً

ويكون مقدر فترة ثقة ١٠٠ (١ − α) ٪ هو :

$$\bar{س} \pm Z_{\frac{1}{2}\alpha} \frac{ع}{\sqrt{ن}}$$

حيث : $\bar{س}$ = ٢٤٢ جنيه ، ع = ٧٣,٢٦٩ جنيها ، ١ − α = ٠,٩٩ ،

α = ٠,٠١ ، $\frac{١}{٢}\alpha$ = ٠,٠٠٥ ، $Z_{٠,٠٠٥}$ = ٢,٥٧٥

∴ تقدير فترة ثقة ٩٩ ٪ للمعلمة μ هو :

$$٢٤٢ \pm ٢,٥٧٥ \frac{٧٣,٢٦٩}{\sqrt{٥٠}}$$

$$٢٤٢ \pm ٢٦,٦٨٢$$

الحد الأدنى لفترة الثقة = ٢١٥,٣١٨ جنيها .

الحد الأعلى لفترة الثقة = ٢٦٨,٦٨٢ جنيها .

وهذا يعني أنه بدرجة ثقة ٩٩ ٪ تحتوي الفترة من ٢١٥,٣١٨ جنيه إلى ٢٦٨,٦٨٢ جنيه على الوسط الحسابي لأجر العمال في المجتمع .

(٢ ـ ٤) تقدير فترة ثقة لمتوسط مجتمع في حالة العينات الصغيرة :

رأينا في المبحث السابق أنه عندما يكون حجم العينة كبيرا (ن ≥ ٣٠) فإننا نستخدم التوزيع المعتدل سواء كانت σ معلومة أم غير معلومة . وذلك لأنه

وفقا لنظرية النهاية لمركزية ، عندما تكون العينات كبيرة فإن التوزيع العينـي للأوساط الحسابية $\bar{س}$ يكون معتدلا تقريبا وذلك بصرف النظـر عـن شـكل التوزيع الأصلي .

وفي كثير من الأحيان يتعذر الحصول على عينة كبيرة سواء بسـبب تكلفتها الباهظة أو بسبب طبيعة التجربة نفسها ، فمثلا يزيد المسـتثمر معرفـة ربح السهم قبل قيامه بعملية الشراء مما يستلزم أراء كثير من بيـوت الخـبرة وما يتطلبه ذلك من تكاليف باهظة مما يجعل المستثمر يكتفي بعينة صغـيرة ، مثال آخر هو اختبار دواء جديد للشفاء من مرض معين ، هنا أيضا يتم استخدام عينة صغيرة بسبب عدم وجود مرضى كثيرين مصابين بهذا المـرض وعلـى استعداد لتجربة الدواء الجديد .

فإذا كان حجم العينة صغيرا (ن < ٣٠) يجب التفرقة بين حالة ما إذا كانت σ معلومة أم لا .

أ ـ إذا كانت σ معلومة :

إذا كانت σ معلومة ، وكان التوزيع الأصلي للمجتمع الذي أخذت منـه العينة هو توزيعا معتدلا ، ففي هذه الحالة يمكننا استخدام التوزيع المعتدل لتقدير فترة الثقة .

وفي هذه الحالة فإن :

$$\bar{س} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{ن}} \qquad (١٠)$$

هو مقدر فترة ثقة ١٠٠ (١ − α) ٪ لمتوسط مجتمع معتدل ومعلوم التباين عندما ن < ٣٠ .

مثال (٥) :

سحبت عينة عشوائية حجمها ٥ مفردات من مجتمع له توزيع معتـدل ع (µ ، ٢^{٢}) وكانت المشاهدات كما يلي :

٢٠ ، ٢٥ ، ٢٨ ، ٣٥ ، ٣٠

والمطلوب : (١) إيجاد فترة ثقة ٩٠ ٪ للوسط الحسابي للمجتمع .

(٢) إيجاد فترة ثقة ٩٥ ٪ للوسط الحسابي للمجتمع .

الحـــل :

(١) ن = ٥ ، σ = ٢

١ − α = ٠,٩٠ ، α = ٠,١٠ ، $\frac{١}{٢}$ α = ٠,٠٥ ؛ $Z_{٠,٠٥}$ = ١,٦٤٥

بما أن المجتمع الذي سحبت منه العينة هو مجتمع معتدل ، σ معلومة ، فإن التوزيع العيني للأوساط الحسابية يتبع التوزيع المعتدل ، ومن ثم :

$$\bar{س} \pm Z_{\frac{١}{٢}\alpha} \frac{\sigma}{\sqrt{ن}}$$

هي مقدر فترة ثقة ١٠٠ (١ − α) ٪ لمتوسط المجتمع µ .

$$\bar{س} = \frac{مجـ س}{ن} = \frac{٢٠ + ٢٥ + ٢٨ + ٣٥ + ٣٠}{٥}$$

= ٢٧,٦

∴ تقدير فترة ثقة ٩٠ ٪ للمعلمة μ هو :

$$٢٧,٦ \pm ١,٦٤٥ \left(\frac{٢}{\sqrt{٥}} \right)$$

$$٢٧,٦ \pm ١,٤٧١$$

الحد الأدنى لفترة الثقة = ٢٦,١٢٩

الحد الأعلى لفترة الثقة = ٢٩,٠٧١

وهذا يعني أنه بدرجة ثقة ٩٠ ٪ تحتوي الفترة من ٢٦,١٢٩ إلى ٢٩,٠٧١ على الوسط الحسابي للمجتمع μ .

(٢) $1 - \alpha = ٠,٩٥$ ، $\alpha = ٠,٠٥$ ، $\frac{١}{٢}\alpha = ٠,٠٢٥$ ، $Z_{٠,٠٢٥} = ١,٩٦$

ويكون تقدير فترة ثقة ٩٥ ٪ لمتوسط المجتمع μ هو :

$$٢٧,٦ \pm ١,٩٦ \left(\frac{٢}{\sqrt{٥}} \right)$$

$$٢٧,٦ \pm ١,٧٥٣$$

الحد الأدنى لفترة الثقة = ٢٥,٨٤٧

الحد الأعلى لفترة الثقة = ٢٩,٣٥٣

وهذا يعني أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٢٥,٨٤٧ إلى ٢٩,٣٥٣ على الوسط الحسابي للمجتمع μ .

ب ـ إذا كانت σ غير معلومة :

وفي كثير من الأحيان تكون σ غير معلومة . فإذا حدث هـذا وكـان التوزيع الذي سحبت منه العينة توزيعا معتدلا أو قريبا من الاعتدال ، نستخدم الانحراف المعياري ع للعينة كمقدر نقطة للمعلمة σ . وفي هـذه الحالـة لا يمكننا استخدام التوزيع المعتدل في تقدير فترة ثقة للمعلمة μ . وهذا لأن المتغير العشوائي :

$$t = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

يتبع توزيع t بدرجات حرية (ن − ١) .

وإذا كانت $t_{(ن - ١ ، \frac{1}{2}\alpha)}$ هي قيمة t التي تجعل إلى يمينها مساحة قدرها $\frac{\alpha}{2}$ فإن :

$$\bar{س} \pm t_{(ن - ١ ، \frac{1}{2}\alpha)} \frac{ع}{\sqrt{ن}} \qquad (١١)$$

هو مقدر فترة ثقة ١٠٠ (١ − α) ٪ لمتوسط مجتمع معتدل أو قريب من الاعتدال وغير معلوم التباين عندما ن < ٣٠ .

مثال (٦) :

أراد أحد المستثمرين تقدير متوسط العائد المتوقع للسهم الذي تصـدره إحدى الشركات ، ولقد قام بالاستعانة بخمس بيوت للخبرة في سـوق الأوراق المالية ، وكانت توقعاتها كالآتي (بالجنيهات) :

١٠,٥ ، ١٦,٢ ، ١١,٤ ، ٩,٦

فإذا علمت أن مجتمع عائد السهم يتوزع توزيعا قريبا من الاعتدال ، فالمطلوب إيجاد فترة ثقة ٩٥ ٪ لمتوسط العائد على السهم في هذه الشركة .

الحـــل :

بما أن التوزيع قريب من الاعتدال ، وغير معلـــوم التبـــاين والعينـــة صغيرة (ن = ٥) فإن :

مقدر فترة ثقة ١٠٠ (١ − α) ٪ لمتوسط المجتمع هو :

$$\bar{س} \pm t_{(ن-1 ، \frac{1}{2}\alpha)} \frac{ع}{\sqrt{ن}}$$

لذلك يجب أولا حساب $\bar{س}$ ، ع .

س	س − $\bar{س}$	(س − $\bar{س}$)^2
١٠,٥	١,٩−	٣,٦١
١٦,٢	٣,٨	١٤,٤٤
١١,٤	١−	١
١٤,٣	١,٩	٣,٦١
٩,٦	٢,٨−	٧,٨٤
٦٢		٣٠,٥

$$\bar{س} = \frac{مجـ س}{ن} = \frac{٦٢}{٥} = ١٢,٤ \text{ جنيها}$$

$$ع^2 = \frac{1}{ن-1} مجـ (س - \bar{س})^2$$

$$= \frac{1}{٥-1} مجـ (س - ١٢,٤)^2$$

= $\frac{١}{٤}$ (٣٠,٥)

= ٧,٦٢٥

ع = ٧,٦٢٥ = ٢,٧٦١ جنيها

١ − α = ٠,٩٥ ، α = ٠,٠٥ ، $\frac{١}{٢}$ α = ٠,٠٢٥ ، ن = ٥

وبالبحث في جدول t نجد أن :

$t_{(٤ ، ٠,٠٢٥)}$ = ٢,٧٧٦

ومن ثم فإن تقدير فترة ثقة ٩٥ ٪ لمتوسط عائد السهم هو :

١٢,٤ ± ٢,٧٧٦ ($\frac{٢,٧٦١}{\sqrt{٥}}$)

١٢,٤ ± ٣,٤٣٨

الحد الأدنى لفترة الثقة = ٨ ، ٢

الحد الأعلى لفترة الثقة = ١٥,٨٢٨

وهذا يعني أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٨,٩٧٢ جنيها إلى ١٥,٨٢٨ على عائد السهم الحقيقي في هذه الشركة .

(٢ ـــ ٥) تقدير فترة ثقة للنسبة في المجتمع :

في كثير من الأحيان يتم تقسيم المجتمع إلى نوعين من المفردات : المفردات التي تتصف بصفة معينة والمفردات التي لا تتصف بهذه الصفة . فمثلا قد ينقسم المجتمع إلى مدخن وغير مدخن ، أو إلى إناث وذكور ، وقد ينقسم مجتمع إنتاج مصنع معين إلى إنتاج معيب وإنتاج غير معيب . فإذا

سحبت مفردة من أحد هذه المجتمعات بطريقة عشوائية ، فإن احتمال تمتع هـذه المفردة بصفة معينة = نسبة هذه الظاهرة في المجتمع = θ . فمثلاً إذا كـــانت نسبة المدخنين في مجتمع معين : θ = ٠,٤٠ ، فإن نسبة غير المدخنين فـــي هذا المجتمع ، (١ − θ) = ١ − ٠,٤٠ = ٠,٦٠ .

وإذا سحبنا عينة عشوائية حجمها ن من هذا المجتمع ، وإذا كان س هو المتغير العشوائي الذي يمثل عدد الذين يتصفون بهذه الصفة في العينة ، فإن س تكون متغيراً عشوائياً له توزيع ذي الحدين .

وفي كثير من الأحيان تكون النسبة في المجتمع (θ) مجهولة ، ونريد تقدير هذه النسبة بنقطة أو بفترة .

أ ـ التقدير بنقطة :

يمكن استخدام النسبة في العينة كمقدر نقطة لنسبة المجتمع θ . فـــإذا استخدمنا الرمز ق للدلالة على النسبة في العينة حيث :

$$\text{ق} = \frac{\text{عدد المفردات التي لها الصفة}}{\text{حجم العينة}} = \frac{\text{س}}{\text{ن}}$$

ويكون مقدر نقطة لنسبة المجتمع θ هو :

$$\hat{\theta} = \text{ق} = \frac{\text{س}}{\text{ن}} \qquad (١٢)$$

ب ـ التقدير بفترة :

$\hat{\theta}$ متغير عشوائي له توزيع احتمالي معين . وسنكتفي في دراستنا هنـــا بأخذ حالة كون حجم العينة كبيراً بحيث تكون : ن $\theta \geq$ ٥ ، ن (١ − θ) $\geq$ ٥ .

وعندما تكون θ غير معلومة ، نستخدم بدلاً مـــن θ مقــدر النقطــة لـــها $\hat{\theta}$ ،

ومن ثم لكي تكون العينة كبيرة يجب أن تكـون ن $\theta \geq 5$ ، ن $(1 - \theta) \geq 5$

وطبقا لنظرية النهاية المركزية ، فمع زيادة حجم العينة يقترب التوزيع العينـــي للنسبة $\hat{\theta}$ من التوزيع المعتدل الذي متوسطه : $\hat{\theta} = \theta$

وتباينه $\sigma^2_{ق} = \dfrac{\theta(1 - \theta)}{ن}$

وبما أننا نريد تقدير θ ، إذن θ غير معلومة ، ومن ثم لا نستطيع حساب تبـاين التوزيع العيني $\sigma^2_{ق}$ ، لذلك نستخدم تباين العينة ع² كمقدر نقطة لتباين التوزيـع العيني $\sigma^2_{ق}$ حيث :

$$ع^2 = \frac{\hat{\theta}(1 - \hat{\theta})}{ن}$$

ومن ثم فإن :

$$\hat{\theta} \pm Z_{\frac{1}{2}\alpha} \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{ن}} \qquad (13)$$

هو مقدر فترة ثقة ١٠٠ $(1 - \alpha)$ ٪ لنسبة مجتمع في حالة العينات الكبيرة حيث : ن $\hat{\theta} \geq 5$ ، ن $(1 - \hat{\theta}) \geq 5$

مثال (٧) :

تريد أحد الشركات القيام بتسويق نوع جديد من مسحوق الغسيل . وقبل القيام بذلك أرادت الشركة القيام بأبحاث للسوق لمعرفة مدى تفضيل الناس لـــهذا المسحوق . فسحبت عينة عشوائية من ٢٠٠ مستهلك وأهدت لهم عبـــوة مجانيـــة ، وبعد استعمالها وجدت الشركة أن ١٤٠ منهم فضلوا هذا المسحوق . المطلوب :

أ ـــ تقدير نقطة لنسبة المستهلكين الذين يفضلون هذا المسحوق .

ب ـــ تقدير فترة ثقة ٩٥ ٪ لنسبة المستهلكين الذين يفضلون هذا المسحوق .

الحـــل :

أ ـ ن = ٢٠٠ ، س = ١٤٠

$\hat{\theta}$ = ق = $\frac{١٤٠}{٢٠٠}$ = ٠,٧ هي مقدر نقطة لنسبة المجتمع θ

ب ـ ١ − α = ٠,٩٥ ، α = ٠,٠٥ ، $\frac{١}{٢}\alpha$ = ٠,٠٢٥ ، $Z_{٠,٠٢٥}$ = ١,٩٦

ن $\hat{\theta}$ = ٢٠٠ (٠,٧) = ١٤٠ > ٥

ن (١ − $\hat{\theta}$) = ٢٠٠ (١ − ٠,٧) = ٦٠ > ٥

∵ ن $\hat{\theta}$ > ٥ ، ن (١ − $\hat{\theta}$) > ٥ ، إذن العينـــة كبـــيرة ،

وبتطبيق نظرية النهاية المركزية نجد أن :

$$\hat{\theta} \pm Z_{\frac{١}{٢}\alpha} \sqrt{\frac{\hat{\theta}\,(١ - \hat{\theta})}{ن}}$$

هو مقدر فترة ثقة ١٠٠ (١ − α) ٪ لنسبة مجتمع .

أي أن :

$$٠,٧ \pm ١,٩٦ \sqrt{\frac{٠,٧\,(٠,٣)}{٢٠٠}}$$

$$٠,٧ \pm ١,٩٦\,(٠,٠٣٢٤)$$

$$٠,٧ \pm ٠,٠٦٣٥$$

الحد الأدنى لفترة الثقة = ٠,٦٣٦٥

الحد الأعلى لفترة الثقة = ٠,٧٣٦٥

أي أنه بدرجة ثقة ٩٥ ٪ تحتوي الفترة من ٠,٦٣٦٥ إلى ٠,٧٦٣٥ على النسبة الحقيقية لتفضيل المستهلكين لهذا المسحوق .

(٢ ــ ٦) تحديد العينة لتقدير متوسط مجتمع :

في الأمثلة السابقة كان حجم العينة معلوم ، ولم نتعرض لسبب اختيار حجم العينة . ولنفرض أننا نريد تقدير بفترة لمتوسط مجتمع ، فما هو حجم العينة الواجب سحبها ؟ فعلى سبيل المثال إذا كان من الممكن الحصول على فترة الثقة التي نريدها من عينة حجمها ٥٠ مفردة ، فعند أخذ عينة حجمها ٢٠٠ مفردة نكون قد أضعنا كثير من النفقات والوقت والجهد بدون مبرر . فإذا كنا نعلم مستوى الثقة وطول فترة الثقة التي نريدها ، يمكننا معرفة حجم العينة التي تعطينا هذه النتائج .

وكما رأينا سلفا فإن مقدر فترة ثقة ١٠٠ (١ − α) ٪ لمتوسط مجتمع هو :

$$\bar{س} \pm Z_{\frac{1}{2}\alpha} \frac{\sigma}{\sqrt{ن}}$$

وبافتراض أن سحب العينة يتم بإرجاع أو أن تكون النسبة بين حجم العينة ن وحجم المجتمع م هي : $\frac{ن}{م} < ٠,٠٥$ ، فإن :

$$Z_{\frac{1}{2}\alpha} \frac{\sigma}{\sqrt{ن}}$$ يسمى بخطأ التقدير (E)

أي أن : $$E = Z_{\frac{1}{2}\alpha} \frac{\sigma}{\sqrt{ن}} \qquad (١٤)$$

ونحصل على قيمة ن بحل معادلة (١٤) نجد أن :

$$ن = \left[\frac{\sigma \, Z_{\frac{1}{2}\alpha}}{E} \right]^{٢} \qquad (١٥)$$

وهنا σ مجهولة ، ويمكن تقديرها باستخدام الانحراف المعياري ع المحسوب من عينة مبدئية ذات حجم صغير متفق عليه ، أو يمكن تقديرها من عينة سبق الحصول عليها من دراسات سابقة أو دراسات مماثلة . وإذا كان توزيع المجتمع معتدلا تقريبا فإن :

المدى = ٦ σ

ومن ثم $\sigma \approx \frac{\text{المدى}}{6} = \frac{\text{أكبر قيمة - أصغر قيمة}}{6}$ (١٦)

ولكن إذا كان سحب العينة يتم بدون إرجاع ، أو عندما تكون النسبة بين حجم العينة ن وحجم المجتمع م هي : $\frac{ن}{م} \geq 0.05$ ، فيجب استخدام معامل التصحيح correction factor كما سبق وبينا في الفصل الأول

حيث : معامل التصحيح = $\frac{م - ن}{م - 1}$ (١٧)

ويصبح خطأ التقدير في هذه الحالة :

$$E = Z_{\frac{1}{2}\alpha} \frac{\sigma}{\sqrt{ن}} \sqrt{\frac{م - ن}{م - 1}} \quad (18)$$

وبحل هذه المعادلة نحصل على قيمة ن ،

$$ن = \frac{م\, Z^2_{\frac{1}{2}\alpha}\, \sigma^2}{E^2 (م - 1) + Z^2_{\frac{1}{2}\alpha}\, \sigma^2} \quad (19)$$

مثال (٨) :

قام قسم البحوث في أحد الشركات بتقدير متوسط الوقت الذي يقضيه العاملين في هذه الشركة للوصول من منازلهم إلى مقر عملهم . فإذا علمت أن قسم البحوث يريد أن يكون التقدير صحيحا في حدود دقيقة واحدة من المتوسط

الأصلي وبدرجة ثقة ٩٩ ٪ ، وأنه بسحب عينة صغيرة مبدئية وجد أن الانحـــراف المعياري يساوي ٥ دقائق ، فالمطلوب : تحديد حجم العينة الواجب سحبها .

الحـــل :

$E = ١$ دقيقة ، $١ - \alpha = ٠,٩٩$ ، $\alpha = ٠,٠١$ ،

$\frac{١}{٢}\alpha = ٠,٠٠٥$ ، $Z_{٠,٠٠٥} = ٢,٥٧٥$ ، $\hat{\sigma} = ع = ٥$ دقائق .

ويكون حجم العينة :

$$ن = \left[\frac{Z_{\frac{١}{٢}\alpha}\ \hat{\sigma}}{E}\right]^{٢}$$

$$= \left[\frac{٢,٥٧٥\ (٥)}{١}\right]^{٢} = ١٦٥,٨ \approx ١٦٦ \text{ عامل}$$

وهذا يعني أنه يجب سحب عينة من ١٦٦ عامل لتقدير متوسط الوقت المنقضي لوصول العاملين من منازلهم إلى مقر عملهم ، وذلك حتى يكون التقدير صحيحا في حدود دقيقة واحدة من المتوسط الأصلي وبدرجة ثقة ٩٩ ٪ .

مثال (٩) :

في المثال السابق ، بالإضافة إلى المعلومات الســابقة . إذا علمــت أن عدد العاملين في هذه الشركة هو ٢٥٠٠ عامل ، المطلوب تحديد حجم العينـــة الواجب سحبها .

الحـــل :

بالإضافة إلى المعلومات السابقة فإن حجم المجتمع م = ٢٥٠٠ عامل

$$\text{ن} = \frac{\text{م}\, Z^{2}_{\frac{1}{2}\alpha}\, \sigma^{2}}{E^{2}\,(\text{م} - 1) + Z^{2}_{\frac{1}{2}\alpha}\, \sigma^{2}}$$

$$= \frac{\text{٢,٥٠٠}\,(\text{٢,٥٧٥})^{2}\,(\text{٥})^{2}}{\text{١}^{2}\,(\text{٢٥٠٠} - \text{١}) + (\text{٢,٥٧٥})^{2}\,(\text{٥})^{2}}$$

$$= \frac{\text{٢,٥٠٠}\,(\text{١٦٥,٧٦٦})}{\text{٢٤٩٩} + \text{١٦٥,٧٦٦}}$$

$$= \frac{\text{٤١٤٤١٤,٠٦٣}}{\text{٢٦٦٤,٧٦٦}}$$

$= \text{١٥٥,٥} \approx \text{١٥٦}$ عامل

(٢ ـ ٧) تحديد حجم العينة لتقدير نسبة المجتمع :

وكما فعلنا بالنسبة لتحديد حجم العينة لتقدير متوسط مجتمع ، سنحدد في هذا المبحث حجم العينة لتقدير نسبة المجتمع . فإذا كانت نسبة حجم العينـــة ن إلى حجم المجتمع م هي $\frac{\text{ن}}{\text{م}} < \text{٠,٠٥}$ ، فإن خطأ التقدير هو :

$$E = Z_{\frac{1}{2}\alpha} \sqrt{\frac{\theta\,(1 - \theta)}{\text{ن}}} \qquad (\text{٢٠})$$

ومنها نجد أن :

$$\text{ن} = \frac{Z^{2}_{\frac{1}{2}\alpha}\, \theta\,(1 - \theta)}{E^{2}} \qquad (\text{٢١})$$

أما إذا كانت نسبة حجم العينة ن إلى حجم المجتمع م هـــي : $\frac{ن}{م} \leq ٠,٠٥$ ، فإن :

$$ن = \frac{م\, Z^{٢}_{\frac{١}{٢}\alpha}\, \theta\,(١ - \theta)}{E^{٢}\,(م - ١) + Z^{٢}_{\frac{١}{٢}\alpha}\, \theta\,(١ - \theta)} \qquad (٢٢)$$

وفي جميع الأحوال استخدمنا هنا θ معلمة المجتمع المجهولة المـــراد تقديرها . وكما فعلنا في المبحث السابق يمكننا أخذ عينة مبدئية صغيرة ومنـــها حساب مقدر نقطة $\hat{\theta}$ للنسبة θ . أو يمكننا استخدام $\theta = ٠,٥$ والتعويض عنـــها في معادلة (٢١) أو (٢٢) ، ولكن هذه الطريقة تجعل حجم العينـــة أكـــبر مـــا يمكن ، هذا لأن ضرب ٠,٥ في ٠,٥ يعطي مقدارا أكبر من ضرب أي نسبتين أخرتين لكل من θ ، $(١ - \theta)$.

مثال (١٠) :

يتعهد أحد المطاعم بتوصيل الطلبات إلى المنازل خلال ٣٠ دقيقة مـــن طلب الطلبية . وأرادت إدارة هذا المطعم تقدير نسبة الطلبيات التي وصلت إلـى المستهلكين خلال ٣٠ دقيقة . فما هو حجم العينة الواجب أخذه حتى يصبح خطأ المعاينة ٠,٠٢ من نسبة المجتمع ، بدرجة ثقة ٩٩ ٪ .

الحـــل :

$E = ٠,٠٢$ ، $١ - \alpha = ٠,٩٩$ ، $\alpha = ٠,٠١$ ،

$\frac{١}{٢}\alpha = ٠,٠٠٥$ ، $Z_{٠,٠٠٥} = ٢,٥٧٥$.

إذن فترة الثقة ٩٩ ٪ هي : $\hat{\theta} \pm ٠,٠٢$

وسنأخذ كتقدير لنسبة المجتمع النسبة $\hat{\theta} = ٠,٥$ ، ومن ثم فإن حجم العينة سيكون :

$$ن = \frac{Z^{٢}_{\frac{١}{٢}\alpha}\ \theta\ (١ - \theta)}{E^{٢}}$$

$$= \frac{(٢,٥٧٥)^{٢}\ (٠,٥)\ (٠,٥)}{(٠,٠٢)^{٢}} = \frac{١,٦٥٨}{٠,٠٠٠٤} = ٤١٤٤ \text{ مستهلك}$$

وهذا يعني أن على مدير المطعم أخذ عينة من ٤١٤٤ مستهلك .

تمارين (٢)

١ ـ في إحدى الشركات تم تدريب عينة عشوائية من ٦٠ موظف على إنجاز عمل معين . ولقد قام فريق من الباحثين بقياس الزمن الذي يستغرقه كل موظف في إنجاز هذا العمل ، فكان متوسط هذا الزمن ١٥ دقيقة بانحراف معياري ٣ دقائق .

والمطلوب :

أ ـ إيجاد تقدير نقطة لمتوسط الزمن المستغرق في هذا العمل .

ب ـ إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط الزمن الذي يتخذه الموظف في إنجاز هذا العمل .

٢ ـ تقوم إحدى الشركات بتعبئة السكر في أكياس من البلاستيك . ولمعرفة متوسط وزن الكيس قامت الشركة بسحب عينة عشوائية من ٥٠ كيس فوجدت أن متوسط وزن الكيس ٩٠٠ جرام بانحراف معياري ٢٥ جرام .

والمطلوب :

أ ـ إيجاد تقدير نقطة لمتوسط وزن الكيس .

ب ـ إيجاد تقدير فترة ثقة ٩٠ ٪ لمتوسط وزن الكيس .

٣ ـ قام قسم البحوث بإحدى شركات الطيران بعمل دراسة لمعرفة عدد المقاعد الشاغرة على رحلات طيرانها . فسحبت عينة عشوائية من ١٠٠ رحلة طيران ، فوجد أن الوسط الحسابي للأماكن الشاغرة بها هي : ١٥,٢ مقعد ، بانحراف معياري ٥,٢ مقعد .

والمطلوب :

أ ـ إيجاد تقدير نقطة لمتوسط المقاعد الشاغرة .

ب ـ إيجاد تقدير فترة ثقة ٩٩ ٪ لمتوسط الأماكن الشاغرة .

٤ ــ أرادت مصلحة البريد معرفة متوسط أرصدة دفاتر البريد في أحد فروعها . فسحبت عينة عشوائية من ٢٥ دفتر توفير فوجدت أن الوسط الحسابي للأرصدة هو ٣٥٠٠ جنيه بانحراف معياري ٤٠٠ جنيه . والمطلوب :

أ ــ إيجاد تقدير نقطة لمتوسط أرصدة دفاتر البريد في هذا الفرع .

ب ــ إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط أرصدة دفاتر البريد في هذا الفرع علما بأن أرصدة دفاتر البريد تتوزع توزيعا معتدلا .

٥ ــ أراد أحد المحلات التجارية الكبيرة معرفة متوسط ما ينفقه العملاء أثناء التسوق في المحل ، لذلك سحبت عينة عشوائية من ٩ عملاء ، فوجد أن ما أنفقه هؤلاء العملاء أثناء التسوق في المحل هي (بالجنيهات) :

١٢٠ ، ١٥٠ ، ١١٥ ، ١٦٥ ، ٧٥ ، ١٨٠ ، ٦٥ ، ١٤٣ ، ٩٧

والمطلوب :

أ ــ إيجاد تقدير نقطة لكل من متوسط وتباين ما ينفقه العملاء أثناء التسوق في هذا المحل .

ب ــ إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط ما ينفقه العملاء أثناء التسوق في هذا المحل ، علما بأن المجتمع الذي سحبت منه العينة يتوزع توزيعا معتدلا .

٦ ــ أراد أحد مديري إنتاج أحد المصانع معرفة أقطار الكرات التي تنتجها أحد الآلات . فقام بسحب عينة عشوائية من ١٠ كرات فوجد أن متوسط قطر الكرات هو ٨٠,٥ ملليمتر بانحراف معياري

٥,٦ ملليمتر . والمطلوب : إيجاد تقدير فترة ثقة ٩٩ ٪ لمتوسط قطر الكرات من إنتاج هذه الآلة علما بأن المجتمع الذي سحبت منه العينة يتوزع توزيعا معتدلا .

٧ ـ لتقدير متوسط عدد حالات الطوارئ التي تصل خلال اليوم الواحد لإحدى المستشفيات ، قام مدير هذه المستشفى بسحب عينة عشوائية من ٢٥ يوما ، فوجد أن الوسط الحسابي = ٢١,٢ حالة والانحراف المعياري ٥,٤ حالة . والمطلوب : إيجاد تقدير فترة ثقة ٩٠ ٪ لمتوسط عدد الحالات التي تصل إلى قسم الطوارئ خلال اليوم الواحد ، علما بإن المجتمع الذي سحبت منه العينة يتوزع توزيعا معتدلا .

٨ ـ أراد أحد مصانع الساعات دراسة دقة نوع معين من الساعات التي ينتجها . فسحب عينة عشوائية من ٧ ساعات من إنتاج المصنع وقام برصد الزمن قبل وبعد ٤٨ ساعة ، فوجد أن عدد الثواني التي قدمتها أو أخرتها الساعة هي على التوالي :

+٣ ، -٥ ، +١٠ ، -٤ ، +٧ ، -٦ ، +٩

والمطلوب :

أ ـ إيجاد تقدير نقطة لكل من متوسط وتباين الزمن الذي تؤخره أو تقدمه هذه الساعات .

ب ـ إيجاد تقدير فترة ثقة ٩٥ ٪ لمتوسط الزمن الذي تؤخره أو تقدمه هذه الساعات علما بأن المجتمع الذي سحبت منه العينة يتوزع توزيعا معتدلا .

٩ ــ لمعرفة نسبة الأمية الثقافية في الجامعة ، سحبت عينة عشــوائية مــن ١٥٠ طالب فوجد أن عدد الأميين ثقافيا هو ٦٥ طالب . فالمطلوب :

أ ــ إيجاد تقدير نقطة لنسبة الأمية الثقافية في الجامعة .

ب ــ إيجاد تقدير فترة ثقة ٩٥ ٪ لنسبة الأمية الثقافية في الجامعة

١٠ ــ قام أحد مراجعي الحسابات بمراجعة حسابات إحدى الشركات ، لذلـك قام بسحب عينة عشوائية من ٢٠٠ مستند فوجد فيها ٢٥ مســتندا بــه أخطاء .

والمطلوب :

أ ــ إيجاد تقدير نقطة لنسبة المستندات التي تحتوي على أخطاء .

ب ــ إيجاد تقدير فترة ثقة ٩٩ ٪ لنسبة المستندات التي تحتوي علــى أخطاء .

١١ ــ تريد إحدى الشركات القيام بعمل دراسة عــن متوســط عــدد أيــام الإجازات المرضية للموظفين بالشركة . فما هو عدد الموظفين الواجب أخذه كعينة لإجراء هذه الدراسة علما بان الشركة تريد أن يكون التقدير صحيحا في حدود ثلاثة أيام من المتوسط الأصلي وبدرجة ثقة ٩٥ ٪ . ولقد سبق لهذه الشركة أن قامت بدراســات ســابقة اســتنتجت منــها أن الانحراف المعياري لعدد أيام الإجازات هو ٩ أيام .

١٢ ــ أرادت إحدى الشركات معرفة نسبة المستهلكين الذين يفضلون نــوع الصابون الجديد الذي طرحته في الأسواق . ولقد قامت الشركة بدراسة سابـة على عينة صغيرة لمعرفة هذه النســبة فوجــدت أنــها ٠,٣٠ . والمطلوب معرفة حجم العينة الواجب أخذه حتى يصبح الخطأ المسموح به ٠,٠١ ، بدرجة ثقة ٩٩ ٪ .

الفصل الثالث

اختبارات الفروض الإحصائية

مقدمة :

لقد رأينا في الفصل السابق كيفية تقدير معلمة المجتمع المجهولة باستخدام إحصائية محسوبة من عينة عشوائية مسحوبة من هذا المجتمع . كما بينا أن هناك طريقتان لهذا التقدير : التقدير بنقطة والتقدير بفترة . وبالنسبة لمستوى معنوية معين ، كلما كانت فترة الثقة أضيق كلما زاد اعتقادنا في أن الإحصائية تمثل تقديراً دقيقاً لمعلمة المجتمع . ومن ناحية أخرى فعندما تكون العينة كبيرة فإن تطبيق نظرية النهاية المركزية تجعلنا لا نكترث بشكل توزيع المجتمع الأصلي ، طالما أن توزيع المعاينة يقترب من التوزيع المعتدل ، ومن ثم يمكننا استخدام القيمة المعيارية Z .

ولكن في كثير من الأحيان نجد أن البعض يدعي أن معلمة المجتمع تساوي قيمة معينة ، فمثلاً قد يدعي مدير إنتاج مصنع المصابيح الكهربائية أن متوسط عمر المصابيح في هذا المصنع هو μ = ١٠٠٠ ساعة . فإذا أخذنا عينة من إنتاج مصابيح هذا المصنع هل يمكننا إثبات ذلك ؟ بالطبع لا ، لأن الطريقة الوحيدة لمعرفة قيمة μ بدقة تتم عن طريق أخذ بيانات عن المجتمع بأسره أي كل إنتاج هذا المصنع في شهر معين مثلاً . ولكن هذه العينة تمكننا من قبول أو رفض إدعاء أن متوسط عمر المصابيح في هذا المصنع هو ١٠٠٠ ساعة . وبما أن العينة ما هي إلا مجموعة من مفردات المجتمع ، لذلك فإن هذا الاستنتاج قد يكون خاطئاً . لإيضاح ذلك كله يقوم هذا الفصل بدراسة اختبارات

الفروض الإحصائية وذلك في أربعة مباحث . يتناول المبحث الأول شرح اختبارات الفروض الإحصائية ، ويتناول المبحث الثاني دراسة اختبارات الفروض الإحصائية المتعلقة بمتوسط مجتمع في حالة العينات الكبيرة ، ويتناول المبحث الثالث دراسة اختبارات الفروض الإحصائية المتعلقة بمتوسط مجتمع في حالة العينات الصغيرة ، ويتناول المبحث الرابع دراسة الفروض الإحصائية المتعلقة بنسبة مجتمع في حالة العينات الكبيرة .

(٣ ـ ١) الفروض الإحصائية :

سبق وذكرنا أنه لمعرفة معلمة المجتمع بدقة يجب إجراء الحصر الشامل . وإذا أخذنا مثال المصابيح الكهربائية ، فإن إدارة إنتاج المصنع أعلنت أن متوسط عمر المصباح من إنتاج هذا المصنع هو μ = ١٠٠٠ ساعة . ولنفرض أننا أخذنا عينة عشوائية من ١٠٠ مصباح من إنتاج هذا المصنع ووجدنا أن متوسط عمر المصباح في هذه العينة هو $\overline{س}$ = ٩٧٠ ساعة . فهل هذا يجعلنا نعتقد أن إدارة إنتاج هذا المصنع تدعي ادعاءا خاطئاً لإيهام المستهلكين بأن عمر المصباح أكبر من عمره الحقيقي ، وأنه في الواقع عمر المصباح أقل من ١٠٠٠ ساعة ؟ لا نستطيع اتهام مصنع المصابيح بهذا الاتهام إلا بعد أن تجري اختباراً للفروض الإحصائية ، لأن هذا الاتهام مبني على معلومات متخذة من عينة عشوائية . وقد يكون الفرق بين متوسط المجتمع ومتوسط العينة ناتج عن خطأ المعاينة فقط ، بمعنى أنه إذا سحبنا عينة عشوائية أخرى من نفس المجتمع فقد نجد أن متوسط عمر المصباح في العينة $\overline{س}$ = ١٠١٠ ساعة مثلاً . لذلك فإننا نقوم بإجراء اختباراً للفروض الإحصائية لمعرفة ما إذا كان الفرق بين متوسط المجتمع μ = ١٠٠٠ ساعة ومتوسط العينة $\overline{س}$ = ٩٧٠ ساعة ناتجا عن الصدفة فقط أم هو فرقاً حقيقياً

وفي الواقع فإن اختبار الفروض الإحصائية يشبه إلى حد كبير الاختبارات العلمية . فالعالم يقوم بوضع صياغة لنظرية معينة ثم بعد ذلك يقوم باختبار هذه النظرية عن طريق المشاهدات . وفي اختبارات الفروض الإحصائية فإن القائم بالبحث الإحصائي يقوم بوضع فرض معين بالنسبة لمعلمة المجتمع ، فهو يفترض أن معلمة المجتمع تساوي قيمة نظرية معينة . ثم بعد ذلك يقوم الباحث بسحب عينة عشوائية من هذا المجتمع ويقوم بمقارنة المشاهدات الناتجة من العينة بالافتراض النظري الذي وضعه : فإذا كانت المشاهدات لا تتفق مع الافتراض النظري فهو يرفض هذا الافتراض ، أما إذا كانت المشاهدات تتفق مع هذا الافتراض ، فإنه يقبله .

وبوجهٍ عام فإن اختبارات الفروض تتضمن أربعة مراحل أساسية وهي :

١ ــ صياغة الفروض الإحصائية .

٢ ــ تعيين إحصائية الاختبار وحسابها .

٣ ــ تحديد مستوى المعنوية والمنطقة الحرجة .

٤ ــ اتخاذ القرار الإحصائي .

وفيما يلي سنقوم بدراسة كل من هذه المراحل على حدة .

١ ــ صياغة الفروض الإحصائية :

لو أخذنا مثال المصابيح الكهربائية فإن افتراض أن ما قاله مدير الإنتاج صحيحاً بأن متوسط عمر المصابيح من إنتاج هذا المصنع يساوي ١٠٠٠ ساعة ، يسمى بفرض العدم Null hypothesis ويرمز له بالرمز H_0 . ويكتب فرض العدم كما يلي :

H_0 : μ = ١٠٠٠ ساعة .

وينص فرض العدم على أن القيمة النظرية لمعلمة المجتمع صحيحة إلى أن يثبت العكس . وبمعنى آخر ، فإن فرض العدم ينص على " عدم وجود فرق " بين معلمة المجتمع وإحصائية العينة ، ومن هنا جاءت تسمية فرض العدم . وفي مثالنا هذا يمكننا كتابة فرض العدم على الصورة الآتية :

$H_0 : \mu \geq$ ١٠٠٠ ساعة .

وهذا لأن عمر المصابيح إذا زاد عن ١٠٠٠ ساعة فهذا أمر مرغوب فيه من وجهة نظر المستهلك ، لذلك فإن كتابة علامة = أو ≤ في فرض العدم لن تؤثر على الاختبار .

ويعتبر هذا المصنع غشاشاً من وجهة نظر المستهلك إذا كان متوسط عمر المصابيح أقل من ١٠٠٠ ساعة .

لذلك فإن الفرض الثاني يسمى بالفرض البديل ويرمز له بالرمز H_1 ويكتب على الصورة التالية :

$H_1 : \mu <$ ١٠٠٠ ساعة :

ويمكن القول بأن الفرض البديل هو الفرض الذي يكون صحيحاً إذا كان فرض العدم غير صحيح .

وفي مثالنا هذا يسمى هذا الاختبار اختبار طرف أيسر a left-tailed test لأن الفرض البديل هنا يحتوي على علامة (<) .

ولنأخذ مثالا آخر ، إذا كانت إحدى الشركات تبيع آلة معينة لإنتاج إحدى قطع الغيار ، ولقد حددت الشركة بأن عدد الوحدات العيبة من إنتاج هذه الآلة في الشهر هو ٩٠ وحدة . هنا يكون فرض العدم :

$H_0 : \mu =$ ٩٠ .

وبما أنه من المستحب أن تكون عدد الوحـــدات المعيبـــة أقـــل مـــن ٩٠ وحدة ، فيمكننا كتابة فرض العدم كما يلي :

$H_0 : \mu \geq$ ٩٠

وبما أنه من غير المستحب (أو غير المرغوب فيه) أن تكون عـــــدد الوحدات المعيبة أكبر من ٩٠ وحدة ، فإن الفرض البديل هو :

$H_1 : \mu <$ ٩٠ وحدة

ويسمى هذا الاختبار اختبار طـــرف أيمـــن a right-tailed test لأن الفرض البديل يحتوي على علامة (<)

ولنأخذ مثالاً آخر ، ولنفرض أن أحد المصانع ينتج نوع معيـــــن مـــن المسامير قطره ٢ ملليمتر . فيكون فرض العدم هو :

$H_0 : \mu =$ ٢ ملليمتر

ومن المستحب في هذه الحالة أن يكون قطر هذه المســـــامير تســـاوي ٢ ملليمتر بالضبط ، ومن غير المستحب (أو المرغوب فيه) أن يزيـــــد هـــذا القطر أو يقل عن ٢ ملليمتر ، لذلك فإن الفرض البديل هو :

$H_1 : \mu \neq$ ٢ ملليمتر

ويسمى مثل هذا الاختبـــــار باختبـــار الطرفيـــن a two-tailed test لأن الفرض البديل يحتوي على علامة (≠) .

وبصفة عامة فإن فرض العدم هو الفرض الذي ينص على أن معلمـــــة المجتمع المعطاة صحيحة ، لذلك يحتوي فرض العدم دائماً على علامة (=) ، وقد يحتوي على إشارة (≥) أو (≤) . أما الفرض البديل فهو الفرض الـــذي ينص على غير المستحب بالنسبة لمعلمة المجتمع ، لذلك فهو لا يحتـــوي أبـــداً على علامة (=) ولكنه يحتوي على علامة (≠) أو (<) أو (>) .

٢ ــ تعيين إحصائية الاختبار وحسابها :

إحصائية الاختبار test statistic هي الإحصائية المستخدمة في اختبارات الفروض الإحصائية . وتتوقف قيمتها على بيانات العينة المسحوبة من المجتمع . ويمكن تعريف إحصائية الاختبار بأنها القاعدة أو المعيار المستخدم لاتخاذ القرار برفض أو قبول فرض العدم . وأغلب إحصائيات الاختبار تكون على الصورة :

$$\text{إحصائية الاختبار} = \frac{\text{إحصائية العينة - القيمة النظرية للمعلمة}}{\text{الخطأ المعياري للتوزيع العيني}}$$

فعلى سبيل المثال إذا أردنا اختبار متوسط مجتمع ، وكان التوزيع العيني يخضع للتوزيع المعتدل ، فإن قيمة Z المعيارية تعتبر إحصائية الاختبار في هذه الحالة ، حيث :

$$Z = \frac{\bar{\text{س}} - \mu}{\frac{\sigma}{\sqrt{\text{ن}}}} \qquad (١)$$

أما إذا كان التوزيع العيني يخضع لتوزيع t ، فإن قيمة t تعتبر إحصائية الاختبار في هذه الحالة ، حيث :

$$t = \frac{\bar{\text{س}} - \mu}{\frac{\text{ع}}{\sqrt{\text{ن}}}} \qquad (٢)$$

٣ ــ تحديد مستوى المعنوية والمنطقة الحرجة :

في نهاية الاختبار يكون لدينا احتمالين : الأول رفض فرض العدم H_0 ، والثاني عدم رفض H_0 . والفرض غير المرفوض قد يكون صحيحاً أو غير صحيح ، كما أن الفرض المرفوض قد يكون صحيحاً أو غير صحيح . ومن ثم فإن لدينا أربع حالات ممكنة عند إجراء أي اختبار إحصائي ألا وهي :

١ ـــ رفض فرض العدم بينما هو غير صحيح .

٢ ـــ رفض فرض العدم بينما هو صحيح .

٣ ـــ عدم رفض فرض العدم بينما هو غير صحيح .

٤ ـــ عدم رفض فرض العدم بينما هو صحيح .

وتعتبر الحالتين (١) ، (٤) مرغوب فيها ، بينما تعتبر الحالات (٢) ، (٣) غير مرغوب فيها وتسمى بالأخطاء . وتسمى حالة رفض فرض العدم بينما هو صحيح بالخطأ من النوع الأول Type 1 error أو خطأ α α error . ففي مثال المصابيح الكهربائية يحدث الخطأ من النوع الأول إذا كان متوسط عمر المصباح في الواقع يساوي ١٠٠٠ ساعة ، ولكن بالصدفة سحبنا عينة عشوائية من هذا المجتمع وكان متوسط عمر المصباح فيها أقل من ١٠٠٠ ساعة ، فقمنا باتخاذ قرار خاطئ برفض فرض العدم H_0 . وتسمى α بمستوى المعنوية Significance level وهي تمثل احتمال حدوث الخطأ من النوع الأول . وبوجه عام يحدد مستوى المعنوية α قبل القيام بالاختبار الإحصائي ، وعادة لا تتعدى قيمة α ٠,١٠ ، وقيم α الأكثر استخداما هي : ٠,٠١ ، ٠,٠٢٥ ، ٠,٠٥ ، ٠,١٠ .

وتسمى حالة عدم رفض فرض العدم بينما هو في الواقع غير صحيح بالخطأ من النوع الثاني type II error أو خطأ β (الحرف الإغريقي β) β error ففي مثال المصابيح الكهربائية يكون متوسط عمر المصباح في المصنع أقل من ١٠٠٠ ساعة ، ولكن نتيجة للصدفة فإن العينة العشوائية المسحوبة من هذا المجتمع كان متوسط عمر المصباح فيها أكبر من أو يساوي ١٠٠٠ ساعة ، ومن ثم فإننا نتخذ قرار خاطئ بعدم رفض H_0 . وتمثل β احتمال حدوث الخطأ من النوع الثاني ، أي احتمال عدم رفض فرض العدم H_0 بينما هو في الواقع غير صحيح . وتسمى القيمة (١ – β) بقوة الاختبار power of the test وهي تمثل احتمال عدم حدوث الخطأ من النوع الثاني . ويبين جدول (١) جميع الحالات الممكنة عند إجراء أي لختبار إحصائي :

القرار \ الواقع	H_0 صحيح	H_0 غير صحيح
رفض H_0	α Type I error	(١ − β) قرار سليم
عدم رفض H_0	(١ − α) قرار سليم	β Type II error

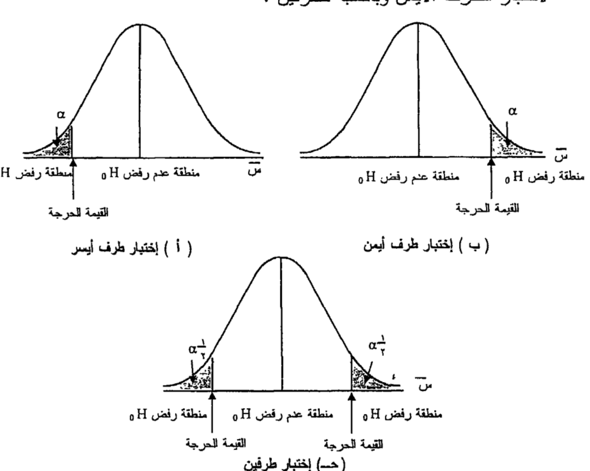
جدول (١) : جميع الحالات الممكنة عند إجراء أي اختبار إحصائي

كلما زادت قيمة α كلما زاد احتمال حدوث الخطأ من النوع الأول أي كلما زاد احتمال رفض فرض العدم بينما هو صحيح . وكلما زادت قيمة β كلما زاد احتمال حدوث الخطأ من النوع الثاني ، أي كلما زاد احتمال عدم رفض فروض العدم بينما هو غير صحيح . ويعتمد كلا من الخطأ من النوع الأول والخطأ من النوع الثاني على الآخر . فبالنسبة لحجم عينة معين ، فلا يمكن تخفيض كلا من α ، β معا في أي اختبار إحصائي : فتخفيض قيمة α يؤدي إلى زيادة قيمة β ، وبالمثل تخفيض قيمة β يؤدي إلى زيادة قيمة α . ولكن هناك طريقة وحيدة يمكن بمقتضاها تخفيض كل من α ، β ألا وهي زيادة حجم العينة .

ويلاحظ في تحليلنا السابق أننا لم نذكر عبارة " قبول فرض العدم " ، لأن كلمة قبول تحمل في طياتها أن هناك قدر كبير من اليقين ، بل ذكرنا بدلا منها عبارة " عدم رفض فرض العدم " .

ولقد سبق وذكرنا عند صياغة الفروض الإحصائية أن هناك ثلاثة أنواع من الاختبارات : اختبار طرف أيسر ، واختبار طرف أيمن ، واختبار طرفين . وعند إجراء أي اختبار من هذه الاختبارات الثلاث نجد أن هناك نقطة على المحور الأفقي للتوزيع العيني تسمى القيمة الحرجة critical value ، وتتحدد

هذه القيمة من جدول التوزيع الاحتمالي لإحصائية الاختبار . وتقوم القيمة الحرجة بتقسيم المحور الأفقي للتوزيع العيني إلى منطقتين : الأولى حول مركز التوزيع تسمى بمنطقة عدم رفض H_0 ، والثانية هي منطقة رفض H_0 ، وهي المنطقة التي تكون مساحتها تحت المنحنى تساوي مستوى المعنوية α . وتسمى منطقة رفض H_0 بالمنطقة الحرجة critical region . ويبين شكل (١) مناطق رفض وعدم رفض H_0 ، بالنسبة لاختبار الطرف الأيسر ، وبالنسبة لاختبار الطرف الأيمن وبالنسبة للطرفين .



شكل (١) : مناطق رفض وعدم رفض H_0 في حالة :

(أ) اختبار طرف أيسر (ب) اختبار طرف أيمن (حـ) اختبار طرفين

٤ ــ اتخاذ القرار الإحصائي :

يمكن اتخاذ القرار الإحصائي بإحدى طريقتين رئيسيتين : الأولى باستخدام القيمة الحرجة والثانية باستخدام مستوى الدلالة المشاهد observed level of significance أي القيمة الاحتمالية لإحصائية الاختبار p-value .

أ ــ باستخدام القيمة الحرجة :

يتم اتخاذ القرار الإحصائي عن طريق مقارنة القيمة المحسوبة لإحصائية الاختبار بالقيمة الحرجة المستخرجة من جدول التوزيع الإحتمالي لإحصائية الاختبار . فإذا كانت القيمة العددية لإحصائية الاختبار تزيد عن القيمة الحرجة ، أي إذا وقعت قيمة إحصائية الاختبار في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى المعنوية α الذي تم تحديده مقدماً . أما إذا كانت القيمة العددية لإحصائية الاختبار تقل عن القيمة الحرجة ، أو بمعنى آخر إذا وقعت إحصائية الاختبار في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض العدم بالمستوى المعنوية α المحدد مقدماً . وإذا كانت إحصائية الاختبار تساوي القيمة الحرجة ، فإننا لا نرفض فرض العدم بمستوى المعنوية α .

وفيما يلي بعض القيم الحرجة الأكثر استخداماً في اختبارات الفروض الإحصائية :

مستوى المعنوية α	نوع الاختبار	القيمة الحرجة Z
٠,٠١	طرف أيمن	$Z_{٠,٠١}$ = ٢,٣٢٦
٠,٠١	طرف أيسر	$-Z_{٠,٠١}$ = −٢,٣٢٦
٠,٠١	طرفين	$Z_{٠,٠٠٥}$ = ٢,٥٧٥
٠,٠٥	طرف أيمن	$Z_{٠,٠٥}$ = ١,٦٤٥
٠,٠٥	طرف أيسر	$-Z_{٠,٠٥}$ = −١,٦٤٥
٠,٠٥	طرفين	$Z_{٠,٠٢٥}$ = ١,٩٦

ب ـ باستخدام القيمة الاحتمالية لإحصائية الاختبار : p- value

لقد سبق وذكرنا أن مستوى المعنوية α يحدد قبل القيام بالاختبار الإحصائي ، وأنه يجب أن يكون صغيراً لا يتعدى ٠,١ . ألا أن اختيار مستوى المعنوية متروك للقائم بالبحث الإحصائي ، فقد يختار أحد الباحثين مستوى معنوية α = ٠,٠٥ ، بينما يفضل باحث آخر اختيار مستوى معنوية α =٠,٠١ . ومن ثم ، باستخدام نفس البيانات فإن أحد الباحثين يستنتج أنه يجب رفض فرض العدم بمستوى معنوية ٠,٠٥ ، بينما يستنتج الباحث الآخر بعدم رفض فرض العدم بمستوى معنوية ٠,٠١ . وحتى مستويات المعنوية المستخدمة ٠,٠١ أو ٠,٠٥ فإنها تحدد بطريقة تحكمية أو بالعادة .

ويمكن تعريف القيمة الاحتمالية p-value لإحصائية اختبار معينة ، بأنها أقل قيمة ممكن أن يأخذها مستوى المعنوية α حتى يتم رفض فرض العدم وفقاً للبيانات المشاهدة . فإذا كانت القيمة الاحتمالية لإحصائية الاختبار كبيرة فإننا لا نرفض فرض العدم ، أما إذا كانت القيمة الاحتمالية لإحصائية الاختبار صغيرة فإننا نرفض فرض العدم .

بالإضافة إلى ذلك يمكن مقارنة القيمة الاحتمالية لإحصائية الاختبار بمستوى المعنوية α المحدد مسبقاً لإجراء الاختبار . فإذا كانت القيمة الاحتمالية لإحصائية الاختبار أكبر من أو تساوي مستوى المعنوية α فإننا لا نرفض فرض العدم H_0 ، وبالعكس إذا كانت القيمة الاحتمالية أقل من مستوى المعنوية α ، فإننا نرفض فرض العدم H_0 .

وخلاصة القول :

لا نرفض فرض العدم H_0 إذا كانت :
قيمة p ≤ مستوى المعنوية α
ونرفض فرض العدم H_0 إذا كانت :
قيمة p > مستوى المعنوية α

ويضع كثير من الباحثين في مقالاتهم المنشورة القيمة الاحتمالية لإحصائية الاختبار عند إجراء الاختبارات الإحصائية ، كما نجدها أيضاً في مخرجات البرامج الجاهزة ، وهذا يعطي القارئ معلومات أكبر من مجرد قبول أو رفض H_0 بالنسبة لمستوى معنوية معين . وبذلك يستطيع القارئ معرفة إلى أي مدى لا تتفق البيانات المشاهدة مع فرض العدم ، بل أكثر من ذلك فإن كل قارئ يستطيع أن يختار قيمة α التي تؤدي إلى رفض فرض العدم . وهذا الأمر لا يتعارض مع اتخاذ القرار الإحصائي باستخدام القيمة الحرجة .(١)

(٣ ـ ٢) اختبارات الفروض الإحصائية المتعلقة بمتوسط مجتمع في حالة العينات الكبيرة :

رأينا في دراستنا لتقدير معالم المجتمع أنه في حالة العينات الكبيرة (ن ≤ ٣٠) ، وفقاً لنظرية النهاية المركزية ، فإن التوزيع العيني للوسط الحسابي $\overline{س}$ يكون معتدلاً تقريباً ، سواء كان تباين المجتمع معلوماً أم لا . لذلك فعندما يكون حجم العينة كبيراً ، فإن المنحنى المعتدل يستخدم في اختبارات الفروض الإحصائية ، ومن ثم فإن إحصائية الاختبار تكون :

$$Z = \frac{\overline{س} - \mu}{\frac{\sigma}{\sqrt{ن}}}$$ إذا كانت σ معلومة

(1) Mendenhall W. . Wackerly D. & Scheaffer R. : "Mathematical Statistics with applications." P. WS-Kent Publishing Co. . Boston, Fourth ed., 1990, pp. 447-450.
Neter J. . Wasserman. W. & Whitmore. "Applied statistics". . Allyn & Bason. Boston. 1993. Pp. 230-234.

$$Z = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$ إذا كانت σ غير معلومة

وخلاصة القول :

في حالة العينات الكبيرة :

إذا كانت σ معلومة : إحصائية الاختبار : $Z = \frac{\bar{س} - \mu}{\frac{\sigma}{\sqrt{ن}}}$

إذا كانت σ غير معلومة : إحصائية الاختبار : $Z = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$ (٣)

مثال (١) :

لنأخذ مثال المصابيح الكهربائية الذي تكلمنا عنه في مستهل هذا الفصل . ففي أحد مصانع إنتاج المصابيح الكهربائية ، أعلنت إدارة الإنتاج أن متوسط عمر المصباح من إنتاج هذا المصنع هو ١٠٠٠ ساعة بانحراف معياري ٩٠ ساعة . ولقد أراد أحد المستوردين شراء شحنة من إنتاج هذا المصنع فأخذ عينة عشوائية من ١٠٠ مصباح وقام بإنارتها جميعا فوجد أن متوسط عمر المصباح $\bar{س}$ = ٩٧٠ ساعة .

والمطلوب :

أولا : هل يجب على هذا المستورد قبول أو رفض الشحنة ؟ استخدم مستوى معنوية ٠,٠١ .

ثانيا : باستخدام القيمة الاحتمالية p ، ما هو قرار المستورد ؟

الحـــل :

μ = ١٠٠٠ ساعة ، σ = ٩٠ ساعة .

$\overline{س}$ = ٩٧٠ ساعة ، ن = ١٠٠ مصباح ، α = ٠,٠١

أولاً :

ويتم الحل على ٤ خطوات :

١ ــ صياغة الفروض الإحصائية :

كما سبق وبينا فإن فرض العدم يفترض أن القيمـــة النظريـــة لمعلمـــة المجتمع صحيحة ، أي أن (H_0 = ١٠٠٠) ولكن بما أنه من المرغوب فيه أن يكون عمر المصباح أكبر من ١٠٠٠ ساعة ، فإن فرض العدم يكون :

H_0 : μ ≤ ١٠٠٠ ساعة .

ويكون الفرض البديل :

H_1 : μ > ١٠٠٠ ساعة .

والاختبار هنا هو اختبار طرف أيسر

٢ ــ تعيين إحصائية الاختبار وحسابها :

بما أن ن = ١٠٠ فإن حجم العينة كبـــير (ن > ٣٠) ، ومـــن ثـــم باستخدام نظرية النهاية المركزية ، فإن التوزيع العيني للأوساط الحسابية يكـــون معتدلاً تقريباً بصرف النظر عن شكل التوزيع الأصلـــي للمجتمـــع . وتكـــون إحصائية الاختبار هي :

$$Z = \frac{\overline{س} - \mu}{\frac{\sigma}{\sqrt{ن}}}$$

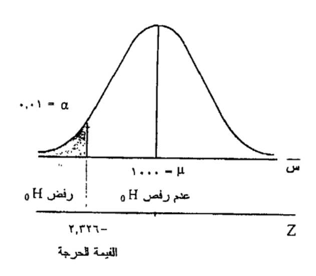
$$Z = \frac{٩٧٠ - ١٠٠٠}{\frac{٩٠}{\sqrt{١٠٠}}} = -٣,٣٣$$

٣ ـ تحديد القيمة الحرجة والمنطقة الحرجة :

α = ٠,٠١ ، $Z_{٠,٠١}$ = ٢,٣٢٦

وبما أن الاختبار هو اختبار طرف أيسر فإن القيمة الحرجة هي Z = −٢,٣٢٦

ويبين شكل (٢) القيمة الحرجة والمنطقة الحرجة أي منطقة رفض H_0 .



شكل (٢)

٤ ـ القرار الإحصائي :

بما أن قيمة Z المحسوبة أقل من القيمة الحرجة :

−٣,٣٣ < −٢,٣٢٦

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠١ ، وهذا يعني أننا نرفض الفرض القائل بأن متوسط عمر المصباح ≤ ١٠٠٠ ساعة ، أو بمعنى آخر فإن الفرق بين متوسط العينة ومتوسط المجتمع فرق معنوي ، بمستوى معنوية ٠,٠١ . ومن ثم يجب رفض الشحنة .

ثانياً : لقد وجدنا أن قيمة Z المحسوبة = −٣,٣٣

نوجد القيمة الاحتمالية p .

بما أن الاختبار طرف أيسر ،

$$\therefore \quad p = ح\,(Z \geq -٣,٣٣) = ١ - \Phi\,(٣,٣٣)$$

$$= ١ - ٠,٩٩٩٥٦٥٨$$

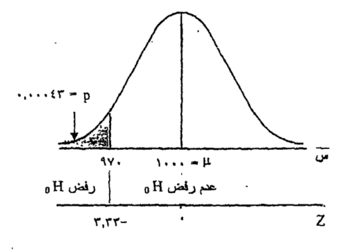
$$= ٠,٠٠٠٤٣٤٢$$

وهنا قيمة p صغيرة جداً = ٠,٠٠٠٤٣٤٢ ، أي أنه يتم رفض فـرض العدم بالنسبة لأي مستوى معنوية أكبر من ٠,٠٠٠٤٣٤٢

وبما أن قيمة p < α المستخدمة في الجزء الأول من هذا المثال .

٠,٠٠٠٤٣ < ٠,٠١ ، إذن نرفض H_0 كما سبق وبينا .

ويبين شكل (٣) قيمة p في هذا المثال .



شكل (٣)

مثال (٢) :

في المثال السابق إذا كان تباين المجتمع غير معلوم ، وكان الانحراف المعياري للعينة ع = ١٠٠ ساعة ، فالمطلوب :

١ ــ هل يجب على هذا المستورد قبول أو رفض هذه الشحنة ؟

٢ ــ باستخدام القيمة الاحتمالية ، استنتج قرار المستورد ؟

الحــــل :

١ ــ μ = ١٠٠٠ ساعة

$\overline{س}$ = ٩٧٠ ساعة ، ع = ١٠٠ ساعة ، ن = ١٠٠ مصباح

الفروض الإحصائية :

$H_0 : \mu \geq$ ١٠٠٠

$H_1 : \mu <$ ١٠٠٠

الاختبار هنا هو اختبار طرف أيسر .

إحصائية الاختبار :

بما أن العينة كبيرة ، فإن التوزيع العيني للأوساط الحسابية يتبع التوزيع المعتدل تقريباً طبقاً لنظرية "النهاية المركزية ، وبما أن σ غير معلومة نستخدم الانحراف المعياري ع للعينة كمقدر نقطة لمعلمة المجتمع σ . وتكون إحصائية الاختبار هي :

$$Z = \frac{\overline{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

$$Z = \frac{٩٧٠ - ١٠٠٠}{\frac{١٠٠}{\sqrt{١٠٠}}} = -٣$$

القيمة الحرجة :

$\alpha = ٠,٠١$ ، $Z_{٠,٠١} = ٢,٣٢٦$

القيمة الحرجة : $Z = -٢,٣٢٦$ لأنه اختبار طرف أيسر ، والقيمة الحرجة والمنطقة الحرجة هنا هي نفسها الموجودة في شكل (٢) .

القرار الإحصائي :

بما أن قيمة Z المحسوبة أقل من القيمة الحرجة :

$-٣ < -٢,٣٢٦$

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠١ . وهذا يعني أن الفرق بين متوسط العينة ومتوسط المجتمع فرق معنوي (أي حقيقي) ولا يمكن إرجاعه للصدفه . ومن ثم فإنه يجب رفض الشحنة .

٢ ــ لقد وجدنا أن قيمة Z المحسوبة = -٣

$\therefore p = ح (Z \leq -٣) = ١ - \Phi (٣)$

$= ١ - ٠,٩٩٨٦٥ = ٠,٠٠١٣٥$

قيمة p صغيرة جدا ، أي أن أقل قيمة ممكنة يمكن أن يأخذها مستوى المعنوية حتى يتم رفض فرض العدم هي ٠,٠٠١٣٥ .

وبما أن : ٠,٠٠١٣٥ < ٠,٠١ ، إذن نرفض H_0 كما سبق واستنتجنا في الجزء الأول من الإجابة .

مثال (٣) :

طبقا لإحصاءات التعداد في أحد المجتمعات وجد أن متوسط دخل الأسرة لا يتعدى ٢٠٠٠٠ دولار ، ولقد قام فريق من الباحثين بأخذ عينة عشوائية من ٥٠ أسرة فوجد أن متوسط دخل الأسرة هو ٢٠٥٠٠ دولار بانحراف معياري ٢٠٠٠ دولار . وبناءا على ذلك استنتج هذا الفريق بأن بيانات التعداد الخاصة بالدخل ليست دقيقة وأن دخل الأسرة أكبر مما جاء في التعداد . هل تتفق مع هذا الفريق في الرأي ؟ استخدم مستوى معنوية ٠,٠١ .

الحـــل :

μ = ٢٠٠٠٠

$\overline{س}$ = ٢٠٥٠٠ ، ع = ٢٠٠٠ ، ن = ٥٠

الفروض الإحصائية :

H_0 : $\mu \leq$ ٢٠٠٠٠

H_1 : $\mu >$ ٢٠٠٠٠

الاختبار هو اختبار طرف أيمن . لأن رأس التمرين هنا يعطي فرض العدم H_0 ، حيث ذكر أن متوسط دخل الأسرة لا يتعدى ٢٠٠٠ دولار (أي $\mu \leq$ ٢٠٠٠٠) ؛ ومن ثم فإن الفرض البديل يكون : H_1 : $\mu >$ ٢٠٠٠٠ ويكون الاختبار هو اختبار طرف أيمن .

إحصائية الاختبار :

بما أن العينة كبيرة ن = ٥٠ > ٣٠ ، فوفقا لنظرية النهاية المركزية فإن توزيع المعاينة للوسط الحسابي $\overline{س}$ يكون معتدلا تقريبا ، ونظرا لأن σ غير

معلومة نستخدم الانحراف المعياري للعينة كمقدر نقطة لها . وتصبح إحصائية الاختبار :

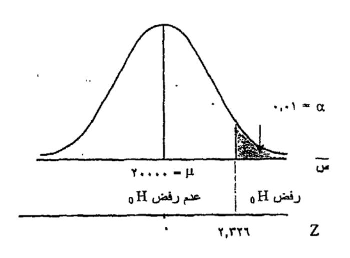
$$Z = \frac{\bar{س} - \mu}{\frac{\sigma}{\sqrt{ن}}}$$

$$Z = \frac{٢٠٥٠٠ - ٢٠٠٠٠}{\frac{٢٠٠٠}{\sqrt{٥٠}}} = ١,٧٦٨$$

تحديد القيمة الحرجة والمنطقة الحرجة :

$\alpha = ٠,٠١$ ، $Z_{٠,٠١}$ = ٢.٣٢٦ = القيمة الحرجة

ويبين شكل (٤) القيمة الحرجة والمنطقة الحرجة .



شكل (٤)

القرار الإحصائي :

بما أن قيمة Z المحسوبة < القيمة الحرجة :

$$١,٧٦٨ < ٢,٣٢٦$$

أي أنها تقع في منطقة عدم رفض H_0 ، ومن ثم فإننا لا نرفض فرض العدم H_0 بمستوى معنوية ٠,٠١ ، أي أنه ليس هناك فرق معنوي بين متوسط المجتمع ومتوسط العينة . ومن ثم فإننا لا نتفق في الرأي مع الفريق بأن دخل الأسرة أكبر مما جاء في التعداد .

مثال (٤) :

ينتج أحد المصانع نوع معين من المسامير قطره ٢ ملليمتر . وأراد أحد التجار شراء شحنة من هذه المسامير فأخذ عينة عشوائية من ٤٠ مسمار فوجد أن الوسط الحسابي لقطر المسمار ١,٨ ملليمتر بانحراف معياري ٠,٣ ملليمتر . ولقد استنتج التاجر بأن هذه المسامير غير مطابقة للمواصفات . فهل تتفق معه في الرأي ؟ استخدم مستوى معنوية ٠,٠٥ .

الحـــل :

μ = ٢ ملليمتر

$\overline{س}$ = ١,٨ ملليمتر ، ع = ٠,٣ ملليمتر ، ن = ٤٠

الفروض الإحصائية :

$H_0 : \mu = ٢$

$H_1 : \mu \neq ٢$

الاختبار هنا اختبار طرفين لأنه بالنسبة للمسامير فإنه من غير المستحب أن يزيد أو يقل قطر المسامير عن ٢ ملليمتر .

إحصائية الاختبار :

العينة هنا = ٤٠ > ٣٠ ، أي عينة كبيرة ، وطبقـــاً لنظريـــة النهايـــة المركزية فإن التوزيع العيني للأوساط الحسابية يكون معتدلاً تقريباً ، ولما كانت σ غير معلومة فإننا نستخدم الانحراف المعياري للعينة ع كمقدر نقطة لـــــها . وتصبح إحصائية الاختبار :

$$Z = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

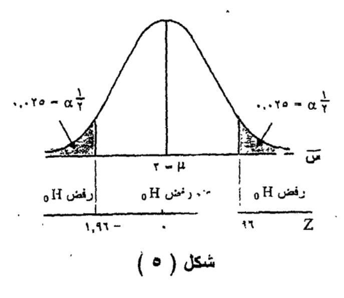
$$Z = \frac{١,٨ - ٢}{\frac{٠,٣}{\sqrt{٤٠}}} = -٤,٢١٦ .$$

تحديد القيمة الحرجة والمنطقة الحرجة :

$\alpha = ٠,٠٥$ ، $\frac{١}{٢}\alpha = ٠,٠٢٥$ ، $Z_{٠,٠٢٥} = ١,٩٦$

بما أن الاختبار هو اختبار طرفين ، لذلك فإن هناك قيمتين حرجتين هما : +١,٩٦ ، -١,٩٦

ويبين شكل (٥) القيمتين الحرجتين والمنطقة الحرجة .



شكل (٥)

القرار الإحصائي :

بما أن قيمة | Z | المحسوبة > القيمة الحرجة :

| ٤,٢١٦ | > ١,٩٦

أي أن قيمة Z تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠٥ . ونتفق مع التاجر في الرأي بأن المسامير غير مطابقة للمواصفات .

العلاقة بين اختبارات الفروض وفترات الثقة :

بالإضافة إلى ما سبق يمكننا إجراء اختبارات الفروض عن طريق فترات الثقة ــ التي قمنا بدراستها في الفصل السابق ــ ولإيضاح ذلك نأخذ مثال (٤) .

مثال (٤) : وهو اختبار طرفين :

H_0 : μ = ٢

H_1 : μ ≠ ٢

وبدلا من إتباع طريقة الحل التي سبق وقمنا بها في حل هذا المثال ، فيمكننا اختبار هذه الفروض بتكوين فترة ثقة ١٠٠ (١ − α) ٪ للمعلمة μ . فإذا كانت هذه الفترة تحتوي على القيمة μ = ٢ فإننا لا ترفض H_0 . وبالعكس إذا كانت هذه الفترة لا تحتوي على القيمة μ = ٢ فإننا ترفض H_0 .

وفي مثالنا هذا بما أن ن كبيرة وطبقا لنظرية النهاية المركزية فإن التوزيع العيني للأوساط الحسابية $\bar{س}$ يتبع التوزيع المعتدل . وهنا مستوى المعنوية α = ٠,٠٥ ، إذن تقدير فترة ثقة ٩٥ ٪ للوسط الحسابي للمعلمة μ هو :

$$\bar{س} \pm Z_{\frac{1}{2}\alpha} \cdot \frac{ع}{\sqrt{ن}}$$

$$١,٨ \pm ١,٩٦ \left(\frac{٠,٣}{\sqrt{٤٠}} \right)$$

$$١,٨ \pm ٠,٠٩٣$$

الحد الأدنى لفترة الثقة = ١,٧٠٧

الحد الأعلى لفترة الثقة = ١,٨٩٣

وبما أن فترة الثقة لا تحتوي على القيمة μ = ٢ ، إذن نرفض فرض العدم بمستوى معنوية ٠,٠٥ وهو نفس القرار الإحصائي الذي توصلنا إليه عند الحل بطريقة القيمة الحرجة .

حساب الخطأ من النوع الثاني β :

عرفنا – في مستهل هذا الفصل – الخطأ من النوع الثاني بأنه احتمال عدم رفض فرض العدم H_0 بينما هو في الواقع غير صحيح . ولحساب الخطأ من النوع الثاني يجب توافر شرطين : الأول أن فرض العدم غير صحيح ، والثاني أن معلمة المجتمع الحقيقية معلومة . وعادة لا نعرف على وجه اليقين إذا كان فرض العدم صحيحا أم لا ، وحتى إذا علمنا أن فرض العدم خاطئ فإننا لا نستطيع معرفة القيمة الحقيقية لمعلمة المجتمع . ومن ثم فإن حساب الخطأ من النوع الثاني لا يكون ممكنا في الواقع العملي .[1] ولإيضاح كيفية حساب الخطأ من النوع الثاني في حالة اختبار طرفين نأخذ المثال التالي .

---

(1) Mann P. S., op. cit. Pp. 474-475.

مثال (٥) :

نفرض أن فرض العدم في مثال (٤) فرض خـــاطئ ، وأن الوسـط الحسابي الحقيقي لقطر المسامير الذي ينتجها هذا المصنع وقت سـحب العينـة العشوائية هو μ = ١,٩٥ ملليمتر . فإذا كان مستوى المعنويـة ٠,٠٥ ، أوجـد احتمال الحصول على الخطأ من النوع الثاني ثم احسب قوة الاختبار .

الحـــل :

ع = ٠,٣ ، ن = ٤٠ ، α = ٠,٠٥ ، $\frac{١}{٢}\alpha$ = ٠,٠٢٥

$Z_{٠,٠٢٥}$ = ١,٩٦

$H_0 : \mu$ = ٢

$H_1 : \mu \neq$ ٢

تتحدد المنطقة الحرجة عندما $| Z | > Z_{\frac{١}{٢}\alpha}$

(١) عندما : $\frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}} > Z_{\frac{١}{٢}\alpha}$ (٤)

أي أن : $\bar{س} > \mu + Z_{\frac{١}{٢}\alpha} \frac{ع}{\sqrt{ن}}$

$\bar{س} >$ ٢ + ١,٩٦ $\frac{٠,٣}{\sqrt{٤٠}}$

$\bar{س} >$ ٢ + ٠,٠٩٣

$\bar{س} >$ ٢,٠٩٣

(٢) وعندما :

$$\frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}} < - Z_{\frac{1}{٢}\alpha} \qquad (٥)$$

$$\bar{س} < \mu - Z_{\frac{1}{٢}\alpha} \frac{ع}{\sqrt{ن}}$$

$$\bar{س} < ٢ - ١,٩٦ \frac{٠,٣}{\sqrt{٤٠}}$$

$$\bar{س} < ٢ - ٠,٠٩٣$$

$$\bar{س} < ١,٩٠٧$$

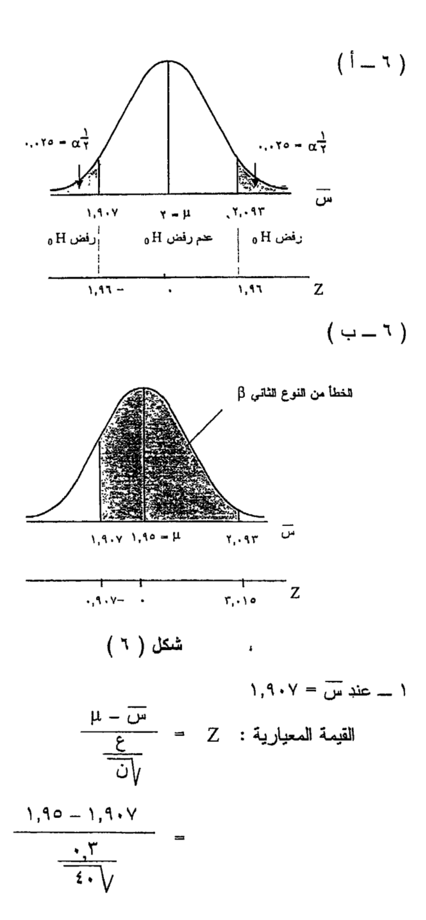
ومن ثم فإن قيمتي $\bar{س}$ المقابلتين للقيمتين الحرجتين −١,٩٦ ، ١,٩٦ هما ١,٩٠٧ ، ٢,٠٩٣ على التوالي كما هو موضح في شكل (٦ ــ أ) . ثم نقوم برسم التوزيع العيني للأوساط الحسابية $\bar{س}$ بأخذ الوسط الحسابي الحقيقي للمجتمع (μ = ١,٩٥) ثم نقوم بحساب الخطأ من النوع الثاني β .

ويبين شكل (٦ ــ أ) التوزيع العيني للأوساط الحسابية عندما كانت معلمة المجتمع μ = ٢ . كما يبين شكل (٦ ــ ب) التوزيع العيني للأوساط الحسابية عندما كانت معلمة المجتمع μ = ١,٩٥ . وتبين المساحة المظللة تحت المنحنى في الشكل (٦ ــ ب) الخطأ من النوع الثاني . ومن الملاحظ أن هذه المساحة تقابل منطقة عدم رفض H_0 على المحور الأفقي في شكل (٦ ــ أ) . ويمكن حساب هذه المساحة ــ أي الخطأ من النوع الثاني β ــ باتباع الخطوات التالية :

١ ــ إيجاد قيمتي Z في شكل (٦ ــ ب) التي تناظر قيمتي $\bar{س}$: ١,٩٠٧ ، ٢,٠٩٣

٢ ــ إيجاد المساحة بين هاتين القيمتين على المنحنى المعتدل المعياري . وقيمة هذه المساحة هي احتمال الحصول على الخطأ من النوع الثاني β .

(٦ ـ أ)



شكل (٦)

١ ـ عند س̄ = ١,٩٠٧

القيمة المعيارية : $Z = \dfrac{\bar{\text{س}} - \mu}{\dfrac{\text{ع}}{\sqrt{\text{ن}}}}$

$$= \dfrac{١,٩٠٧ - ١,٩٥}{\dfrac{٠,٣}{\sqrt{٤٠}}} = ٠,٩٠٧-$$

عند $\bar{س}$ = ٢,٠٩٣

القيمة المعيارية : $Z = \frac{٢,٠٩٣ - ١,٩٥}{\frac{٠,٣}{\sqrt{٤٠}}} = ٣,٠١٥$

β = ح (-٠,٩٠٧ < Z < ٣,٠١٥)

= Φ (٣,٠١٥) – Φ (-٠,٩٠٧)

= Φ (٣,٠١٥) – [١ – Φ (٠,٩٠٧)]

= ٠,٩٩٨٧١٥ – (١ – ٠,٨١٨٦)

= ٠,٨١٧٣١٥

وهذا يعني أن احتمال عدم رفض فرض العدم بينما هو في الواقع غير صحيح = ٠,٨١٧٣١٥

قوة الاختبار = ١ – β

= ١ – ٠,٨١٧٣١٥

= ٠,١٨٢٦٨٥

ومن ثم فإن احتمال رفض فرض العدم بينما هو في الواقع غير صحيح هو ٠,١٨٢٦٨٥ .

مثال (٦) :

ولنفرض في مثال (٣) أن فرض العدم غير صحيح وأن متوسط دخل الأسرة في المجتمع هو ٢٠٣٠٠ دولار ، فإذا كان مستوى المعنوية ٠,٠١ أوجد احتمال الحصول على الخطأ من النوع الثاني ثم احسب قوة الاختبار .

الحـــل :

ع = ٢٠٠٠٠ ، ن = ٥٠ ، α = ٠,٠١ ، $Z_{٠,٠١}$ = ٢,٣٢٦

$H_0 : \mu \leq$ ٢٠٠٠٠

$H_1 : \mu >$ ٢٠٠٠٠

الاختبار هنا هو اختبار طرف أيمن .

لذلك فإن المنطقة الحرجة تكون عندما $Z > Z_\alpha$

أي :
$$\frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}} > Z_\alpha$$

$$\bar{س} > \mu + Z_\alpha \frac{ع}{\sqrt{ن}}$$

$$\bar{س} > ٢٠٠٠٠ + ٢,٣٢٦ \frac{٢٠٠٠}{\sqrt{٥٠}}$$

$$\bar{س} > ٢٠٦٥٧,٨٩٢$$

إذن قيمة $\bar{س}$ المقابلة للقيمة الحرجة $Z_{٠,٠١}$ = ٢,٣٢٦ هي $\bar{س}$ = ٢٠٦٥٧,٨٩٢

ويبين شكل (٧ ـ أ) التوزيع العيني للأوساط الحسابية عندما كانت معلمة المجتمع μ = ٢٠٠٠٠ ، مبينا القيمة الحرجة والقيمة المقابلة لها من قيم $\bar{س}$.

نوجد قيمة Z التي تناظر قيمة $\bar{س}$ = ٢٠٦٧,٨٩٢ بالنسبة للتوزيع العيني الجديد ، حيث μ = ٢٠٣٠٠ كما هو مبين في شكل (٧ ـ ب) .

$$Z = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

$$= \frac{٢٠٦٥٧,٨٩٢ - ٢٠٣٠٠}{\frac{٢٠٠٠}{\sqrt{٥٠}}}$$

$$= ١,٢٦٥$$

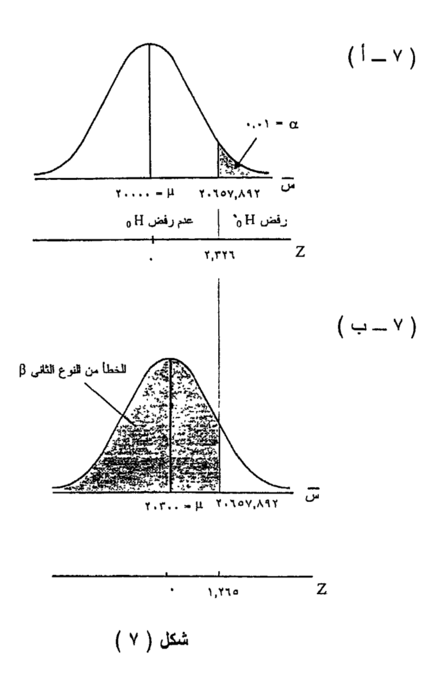
ويمكن حساب الخطأ من النوع الثاني β بأنه المساحة على يسار قيمة Z = ١,٢٦٥ تحت منحنى التوزيع العيني الجديد في شكل (٧ ـ ب) .

أي أن :

$$\beta = ح (Z < ١,٢٦٥) = \Phi (١,٢٦٥)$$

$$= ٠,٨٩٧١$$

أي أن احتمال عدم رفض فرض العدم بينما هو غير صحيح = ٠,٨٩٧١



شكل (٧)

وتكون قوة الاختبار = ١ − β

= ١ − ٠,٨٩٧١ = ٠,١٠٢٩

أي أن احتمال رفض H_0 بينما هو غير صحيح = ٠,١٠٢٩

(٣ ــ ٣) اختبارات الفروض الإحصائية المتعلقة بمتوسط مجتمع في حالة العينات الصغيرة :

كما رأينا سلفاً في الفصل السابق ــ في مبحث (٢ ــ ٤) عند دراسة تقدير فترة ثقة لمتوسط مجتمع في حالة العينات الصغيرة ــ إذا كانت العينات صغيرة (ن < ٣٠) يجب التفرقة بين حالة ما إذا كانت σ معلومة أم غير معلومة .

أ ــ **إذا كانت σ معلومة** : وكان التوزيع الأصلي للمجتمع الذي أخذت منه العينة هو توزيعاً معتدلاً أو قريباً من الاعتدال ، فإن التوزيع العيني للأوساط الحسابية يكون أيضاً معتدلاً ، ومن ثم فإن إحصائية الاختيار في هذه الحالة هي :

$$Z = \frac{\bar{\text{س}} - \mu}{\frac{\sigma}{\sqrt{\text{ن}}}}$$

ب ــ **إذا كانت σ غير معلومة** : وكان التوزيع الذي سحبت منه العينة توزيعاً معتدلاً أو قريباً من الاعتدال ، فإننا نستخدم الانحراف المعياري للعينة ع كمقدر نقطة للمعلمة σ ، فإن إحصائية الاختيار في هذه الحالة هي :

$$t = \frac{\bar{\text{س}} - \mu}{\frac{\text{ع}}{\sqrt{\text{ن}}}}$$

وهي تتبع توزيع t .

وخلاصة القول :

إذا كانت σ معلومة : إحصائية الاختبار : $Z = \frac{\bar{س} - \mu}{\frac{\sigma}{\sqrt{ن}}}$

(٦)

إذا كانت σ غير معلومة : إحصائية الاختبار : $t = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$

ولإجراء اختبارات الفروض نتبع نفس الطريقة التي اتبعناها عند إجراء اختبارات الفروض في حالة العينات الكبيرة مع اختلاف واحد ألا وهو استخدام إحصائية الاختبار t بدلاً من Z .

مثال (٧) :

ينتج أحد مصانع لعب الأطفال لعبة جديدة تربوية عبارة عن مكعبات يقوم الأطفال بتجميعها بأسرع وقت ممكن . ولقد أعلن هذا المصنع أن متوسط الوقت الذي يقضيه طفل عمره ٥ سنوات لتجميع هذه المكعبات هو ٢٥ دقيقة على الأقل . ويتبع توزيع وقت تجميع المكعبات التوزيع المعتدل . ولقد ادعت إحدى المدارس الخاصة بأن تلاميذها الذين يبلغون من العمر ٥ سنوات متميزون ويتمتعون بقدرات خاصة وأن الوقت الذي يقضونه في تجميع هذه المكعبات أقل من ٢٥ دقيقة . ولقد أخذت هذه المدرسة عينة عشوائية من ٢٠ تلميذ من فئة عمر ٥ سنوات ، وكان متوسط الوقت المنقضي في تجميع هذه المكعبات هو ٢٢ دقيقة بانحراف معياري ٢,٥ دقيقة . فهل تتفق مع رأي هذه المدرسة ؟ استخدم مستوى معنوية α = ٠,٠١ .

الحـــل :

μ = ٢٥ دقيقة ، $\bar{س}$ = ٢٢ دقيقة ، ع = ٢,٥ دقيقة

ن = ٢٠ ، α = ٠,٠١

الفروض الإحصائية :

H_0 : μ ≥ ٢٥

H_1 : μ < ٢٥

وهو اختبار طرف أيسر ، لأن رأس التمرين ينص على أن متوسط الوقت الذي يقضيه الطفل لتجميع المكعبات هو ٢٥ دقيقة على الأقل أي أن H_0 : μ ≥ ٢٥ ، ويكون H_1 : μ < ٢٥ .

إحصائية الاختبار :

العينة هنا صغيرة (ن < ٣٠) ، والمجتمع الذي سحبت منـــه العينـــة يتوزع توزيعاً معتدلاً ، والانحراف المعياري للمجتمع غير معلوم ، لذلك فـــإن التوزيع العيني للأوساط الحسابية $\bar{س}$ يخضع لتوزيع t ، . وتكـــون إحصـائيـــة الاختبار هي :

$$t = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

$$= \frac{٢٢ - ٢٥}{\frac{٢,٥}{\sqrt{٢٠}}} = -٥,٣٦٦٦$$

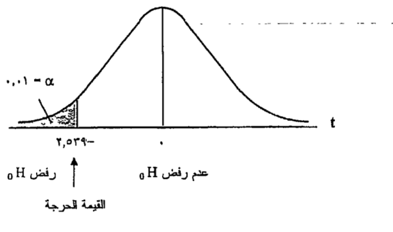
تحديد القيمة الحرجة والمنطقة الحرجة :

α = ٠,٠١ ، ن – ١ = ١٩

القيمة الحرجة = – $t_{(19 ، 0.01)}$ = –٢,٥٣٩

وتكون المنطقة الحرجة حيث : $t < -t_{(19 ، 0.01)}$

كما هو مبين في شكل (٨)



شكل (٨)

القرار الإحصائي :

بما أن قيمة t المحسوبة أقل من القيمة الحرجة :

–٥,٣٦٦٦ > –٢,٥٣٩

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض فرض العدم بمستوى معنوية ٠,٠١ . وهذا يعني أننا نتفق مع ادعاء المدرسة بأن تلاميذها تأخذ وقت أقل في جميع المكعبات ، وهذا بمستوى معنوية ٠,٠١ .

مثال (٨) :

أعلن أحد مديري شركات الطيران أن متوسط عدد الأماكن الشاغرة على أحد خطوط رحلات الطيران لا يتعدى ١٠ مقاعد . ولمعرفة مدى صحة هذا الإدعاء قامت إدارة البحوث باختيار عينة عشوائية من ١٦ رحلة طيران على هذا الخط فوجد أن الوسط الحسابي ١٢,٧ مقعد والانحراف المعياري ٤,٥ مقعد . فإذا علمت أن توزيع الأماكن الشاغرة في الطائرات يتبع التوزيع المعتدل ، فهل تتفق مع مدير شركة الطيران في الرأي ؟ استخدم مستوى معنوية ٠,٠٥ .

الحـــل :

μ = ١٠ ، ن = ١٦ ، $\overline{س}$ = ١٢,٧ ، ع = ٤,٥ ، α = ٠,٠٥

الفروض الإحصائية :

$H_0 : \mu \leq$ ١٠

$H_1 : \mu >$ ١٠

هذا الاختبار هو اختبار طرف أيمن ، لأن رأس التمرين ينص على أن عدد الأماكن الشاغرة لا يتعدى ١٠ مقاعد أي أن $H_0 : \mu \leq$ ١٠ ، ويكون $H_1 : \mu >$ ١٠ .

إحصائية الاختبار :

العينة هنا صغيرة (ن < ٣٠) ، والتوزيع الذي سحبت منه العينة توزيعاً معتدلاً ، والانحراف المعياري للمجتمع غير معلوم ، لذلك فإن التوزيع العيني للأوساط الحسابية $\overline{س}$ يخضع لتوزيع t ، وتكون إحصائية الاختبار هي :

$$t = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

$$= \frac{١٢,٧ - ١٠}{\frac{٤,٥}{\sqrt{١٦}}} = ٢,٤$$

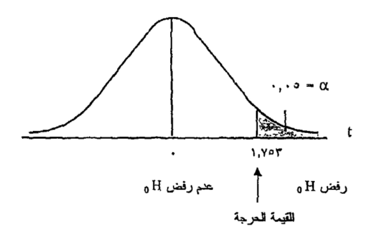
تحديد القيمة الحرجة والمنطقة الحرجة :

α = ٠,٠٥ ، ن − ١ = ١٥

القيمة الحرجة = $t_{(١٥ ، ٠,٠٥)}$ = ١,٧٥٣

وتكون المنطقة الحرجة حيث : $t > t_{(١٥ ، ٠,٠٥)}$

كما هو مبين في شكل (٩)



شكل (٩)

القرار الإحصائي :

بما أن قيمة t المحسوبة أكبر من القيمة الحرجة :

٢,٤ > ١,٧٥٣

أي أنها تقع في منطقة رفض H_0 ، فإننا نرفض H_0 بمستوى معنوية ٠,٠٥ . وهذا يعني أننا نرفض ادعاء المدير بأن متوسط عدد الأماكن الشاغرة على هذا الخط من الطيران لا يتعدى ١٠ مقاعد .

مثال (٩) :

في مشروع التخرج قام خمسة من الطلاب بقياس مساحة قطعة أرض . وكانت المساحة حسب مقاس كل منهم بالفدان هي :

٥,٢٦ ، ٥,٢٣ ، ٥,٢١ ، ٥,٢٧ ، ٥,٢٢

فإذا علمت أن توزيع قياسات مساحة الأرض هو توزيع معتدل ، وأن مالك الأرض أعلن أن المساحة الحقيقية لهذه الأرض هي ٥,٢٢ فدان ، المطلوب : اختبار الفرض القائل بأن المساحة الحقيقية لقطعة الأرض هي ٥,٢٢ فدان ، استخدم مستوى معنوية ٠.٠٥ .

الحل :

μ = ٥,٢٢ فدان ، α = ٠,٠٥

من بيانات العينة نحسب أولا الوسط الحسابي $\bar{س}$ والانحراف المعياري ع .

المساحة (بالفدان) س	س − $\bar{س}$	$(س - \bar{س})^2$
٥,٢١	−٠,٠٢٨	٠,٠٠٠٧٨٤
٥,٢٢	−٠,٠١٨	٠,٠٠٠٣٢٤
٥,٢٣	−٠,٠٠٨	٠,٠٠٠٠٦٤
٥,٢٦	٠,٠٢٢	٠,٠٠٠٤٨٤
٥,٢٧	٠,٠٣٢	٠,٠٠١٠٢٤
٢٦,١٩		٠,٠٠٢٦٨٠

$$\bar{س} = \frac{\text{مجـ س}}{ن} = \frac{٢٦,١٩}{٥} = ٥,٢٣٨ \text{ فدان}$$

$$ع^2 = \frac{١}{ن - ١} \text{ مجـ } (س - \bar{س})^2$$

$$= \frac{1}{5-1} (0.00268) = 0.00067$$

$$ع = \sqrt{0.00067} = 0.026 \text{ فدان}$$

الفروض الإحصائية :

$H_0 : \mu = 5.22$ ،

$H_1 : \mu \neq 5.22$

والاختبار هنا اختبار طرفين .

إحصائية الاختيار :

العينة هنا صغيرة (ن = ٥ < ٣٠) ، المجتمع الذي سحبت منه هذه العينة يتبع التوزيع المعتدل ، والانحراف المعياري للمجتمع غير معلوم ، لذلك فإن التوزيع العيني للأوساط الحسابية $\bar{س}$ يخضع لتوزيع t ، وتكون إحصائية الاختبار هي :

$$t = \frac{\bar{س} - \mu}{\frac{ع}{\sqrt{ن}}}$$

$$= \frac{5.238 - 5.22}{\frac{0.026}{\sqrt{5}}} = 1.548$$

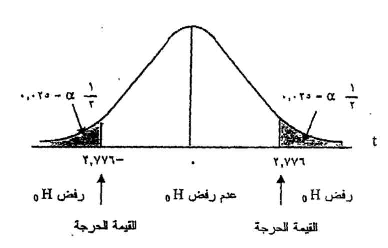
تحديد القيمة الحرجة والمنطقة الحرجة :

$\alpha = 0.05$ ، $\frac{1}{2}\alpha = 0.025$ ، ن − ١ = ٤

بما أن الاختبار هو اختبار طرفين ، إذن هناك قيمتين حرجتين وهما :

$$\pm t_{(4\,،\,0.025)} = \pm 2.776$$

وتكون المنطقة الحرجة حيث : $|t| > t_{(4\,،\,0.025)}$ كما هو مبين في شكل (١٠)



شكل (١٠)

القرار الإحصائي :

بما أن قيمة $|t|$ < القيمة الحرجة

$$|1.548| < 2.776$$

أي أن t تقع في منطقة عدم رفض H_0 ، إذن لا نرفض فرض العدم بمستوى معنوية ٠,٠١ . أي أننا لا نرفض إعلان مالك الأرض بأن المساحة الحقيقية لقطعة الأرض هي ٥,٢٢ فدان .

(٣ ــ ٤) اختبارات الفروض الإحصائية المتعلقة بنسبة مجتمع في حالة العينات الكبيرة :

سبق وتناولنا في المبحث (٢ ــ ٥) تقدير فترة الثقة للنسبة في المجتمع حيث اقتصرت دراستنا علي العينات الكبيرة حيث ن θ ≤ ٥ . وطبقا لنظرية النهاية المركزية فمع زيادة حجم العينة فإن التوزيع العيني يقترب من التوزيع المعتدل الذي متوسطه = θ ، وتباينه = $\frac{\theta(1-\theta)}{ن}$.

وبالنسبة لاختبارات الفروض الإحصائية المتعلقة بنسبة مجتمع في حالة العينات الكبيرة ، فإنها تشبه إلى حد كبير اختبارات الفروض المتعلقة بمتوسط مجتمع في حالة العينات الكبيرة . لذلك سنتبع نفس المراحل الأربعة السابق استخدامها في اختبارات الفروض . فبالنسبة لصياغة الفروض الإحصائية المتعلقة بالنسبة θ ، بافتراض أن θ_0 هي قيمة معلمة المجتمع θ :

فعند اختبار طرف أيسر :

$H_0 : \theta \geq \theta_0$

$H_1 : \theta < \theta_0$

فعند اختبار طرف أيمن :

$H_0 : \theta \leq \theta_0$

$H_1 : \theta > \theta_0$

وعند اختبار طرفين :

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

أما بالنسبة لإحصائية الاختبار فستكون :

$$Z = \frac{ق - \theta_0}{\sqrt{\frac{\theta_0 (١ - \theta_0)}{ن}}} \quad (٧)$$

حيث : θ = قيمة النسبة في المجتمع .

ق = $\hat{\theta}$ = النسبة في العينة

ن = حجم العينة

ويتم تحديد القيمة الحرجة من جدول المنحنى المعتدل المعياري طبقا لمستوى المعنوية α ، فتكون $-Z_\alpha$ في حالة اختبار الطرف الأيسر ، Z_α في حالة اختبار الطرف الأيمن ، $Z_{\frac{١}{٢}\alpha}$ في حالة اختبار الطرفين . ويتم تحديد المنطقة الحرجة بنفس الطريقة التي اتبعناها من قبل ، وكذلك فإن القرار الإحصائي يتم أيضا بنفس الطريقة السابق اتباعها . ويتضح ذلك من الأمثلة التالية .

مثال (١٠) :

أعلن أحد المسئولين أن نسبة الأمية في أحد المجتمعات الريفية لا تتعدى ٤٠ ٪ . ولقد قام فريق من الباحثين بسحب عينة عشوائية من ٢٠٠ فرد من هذا المجتمع فوجد أن عدد الأميين فيها ٩٠ شخص ، فهل تؤيد قول المسئول بأن نسبة الأمية لا تتعدى ٤٠ ٪ ؟ استخدم مستوى معنوية ٠,٠٥ .

الحـــل :

θ_0 = ٠,٤٠ ، ن = ٢٠٠ ، ق = $\frac{٩٠}{٢٠٠}$ = ٠,٤٥ .

الفروض الإحصائية :

$H_0 : \theta \leq$ ٠,٤٠

$H_1 : \theta >$ ٠,٤٠

الاختبار هنا اختبار طرف أيمن

إحصائية الاختبار :

بما أن العينة كبيرة : $ن\ \theta = ٢٠٠\ (٠,٤٠) = ٨٠ > ٥$

$ن\ (١ - \theta) = ٢٠٠\ (٠,٦) = ١٢٠ > ٥$

فيمكننا استخدام التوزيع المعتدل ، طبقاً لنظرية النهايـــة المركزيـــة . وتكـــون إحصائية الاختبار هي :

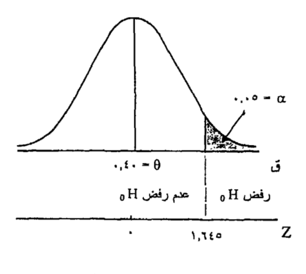
$$Z = \frac{ق - \theta_0}{\sqrt{\frac{\theta_0\ (١ - \theta_0)}{ن}}}$$

$$= \frac{٠,٤٥ - ٠,٤٠}{\sqrt{\frac{٠,٤٠\ (١ - ٠,٤٠)}{٢٠٠}}} = ١,٤٤٣$$

تحديد القيمة الحرجة والمنطقة الحرجة :

$\alpha = ٠,٠٥$ ، $Z_{٠,٠٥} = ١.٦٤٥ =$ القيمة الحرجة .

ويبين شكل (١١) القيمة الحرجة والمنطقة الحرجة .



شكل (١١)

القرار الإحصائي :

بما أن قيمة Z المحسوبة < القيمة الحرجة :

$$١,٦٤٥ > ١,٤٤٣$$

أي أنها تقع في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض العدم القائل بـأن نسبة الأمية لا تتعدى ٠,٤٠ ، بمستوى معنوية α = ٠,٠٥ . أي أننا نؤيد قول المسئول بأن نسبة الأمية لا تتعدى ٤٠ ٪.

مثال (١١) :

أرادت إحدى الشركات طرح منتج جديد في الأسواق . ولقد قرر مدير هذه الشركة أنه سيقوم بإنزال هذا المنتج في الأسواق إذا كان هناك على الأقل ٣٣ ٪ من المستهلكين يفضلونه . وتريد الشركة معرفة نسبة المستهلكين الذين يفضلون هذا المنتج ، لذلك قام قسم الأبحاث بالشركة بسحب عينة عشوائية من ٣٠٠ مستهلك وأعطى لهم المنتج الجديد مجاناً . وبعد تجريبه تبين أن ٨٧ منهم يفضلون هذا المنتج . فهل يجب على الشركة تسويق هذا المنتج أم لا ؟ استخدم مستوى معنوية ٠,٠٥ .

الحل :

θ_0 = ٠,٣٣ ، ن = ٣٠٠ ، ق = $\frac{٨٧}{٣٠٠}$ = ٠,٢٩

الفروض الإحصائية :

$H_0 : \theta \leq$ ٠,٣٣

$H_1 : \theta >$ ٠,٣٣

والاختبار هنا هو اختبار طرف أيسر .

إحصائية الاختبار :

بما أن العينة كبيرة : ن θ = ٣٠٠ (٠,٣٣) = ٩٩ > ٥

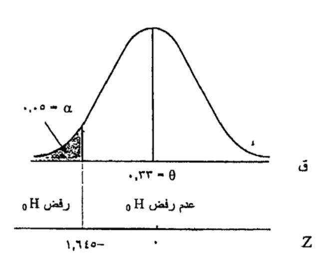
ن (١ − θ) = ٣٠٠ (٠,٦٧) = ٢٠١ > ٥

فطبقاً لنظرية النهاية المركزية يمكننا استخدام التوزيع المعتدل . وتكون إحصائية الاختبار هي :

$$Z = \frac{ق - \theta_0}{\sqrt{\frac{\theta_0 (١ - \theta_0)}{ن}}}$$

$$= \frac{٠,٢٩ - ٠,٣٣}{\sqrt{\frac{٠,٣٣ (٠,٦٧)}{٣٠٠}}} = -١,٤٧٣$$

تحديد القيمة الحرجة والمنطقة الحرجة :



شكل (١٢)

α = ٠,٠٥ ، $Z_{٠,٠٥}$ = ١,٦٤٥

القيمة الحرجة هنا = −١,٦٤٥ لأن الاختبار طرف أيسر ، ويبين شكل (١٢) القيمة الحرجة والمنطقة الحرجة .

القرار الإحصائي :

بما أن قيمة Z المحسوبة أكبر من القيمة الحرجة :

−١,٤٧٣ > −١,٦٤٥

أي أنها تقع في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض العدم بمستوى معنوية α = ٠,٠٥ . وهذا يعني أننا لا نرفض الفرض القائل بأن ٠,٣٣ من المستهلكين يفضلون هذا المنتج ، ومن ثم فإن على الشركة القيام بتسويق هذا المنتج .

مثال (١٢) :

أعلنت أحد المجلات أن نسبة المدخنين بين طلبة الجامعة هي ٣٨ ٪ . ولقد قام فريق من الباحثين بسحب عينة عشوائية من ٤٠٠ طالب فوجد أن عدد المدخنين ١٦٦ طالب . اختبر الفرض القائل بأن نسبة المدخنين تختلف عن ٣٨ ٪ بمستوى معنوية ٠,٠١ .

الحـــل :

θ_0 = ٠,٣٨ ، ن = ٤٠٠ ، ق = $\frac{١٦٥}{٤٠٠}$ = ٠,٤١٥

الفروض الإحصائية :

H_0 : θ = ٠,٣٨

H_1 : $\theta \neq$ ٠,٣٨

والاختبار هنا هو اختبار طرفين .

إحصائية الاختبار :

بما أن العينة كبيرة : ن θ = ٤٠٠ (٠,٣٨) = ١٥٢ > ٥

ن (١ − θ) = ٤٠٠ (٠,٦٢) = ٢٤٠ > ٥

وطبقاً لنظرية النهاية المركزية يمكننا استخدام التوزيع المعتدل وتكون إحصائية الاختبار هي :

$$Z = \frac{\text{ق} - \theta_0}{\sqrt{\frac{\theta_0 \left(\text{١} - \theta_0 \right)}{\text{ن}}}}$$

$$= \frac{\text{٠,٤١٥} - \text{٠,٣٨}}{\sqrt{\frac{\text{٠,٣٨} \left(\text{٠,٦٢} \right)}{\text{٤٠٠}}}} = \text{١,٤٤٢}$$

تحديد القيمة الحرجة والمنطقة الحرجة :

α = ٠,٠١ ، $\frac{\text{١}}{\text{٢}}\alpha$ = ٠,٠٠٥ ، $Z_{\text{٠,٠٠٥}}$ = ٢,٥٧٥

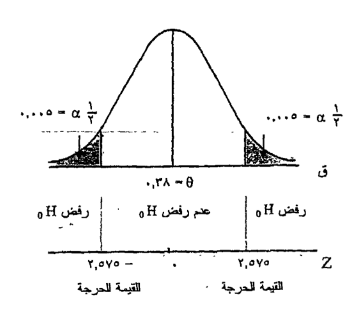
بما أن الاختبار هو اختبار طرفين ، إذن هناك قيمتين حرجتين همـا : +٢,٥٧٥ ، −٢,٥٧٥ . ويبين شكل (١٣) القيمتين الحرجتين والمنطقة الحرجة .

القرار الإحصائي :

بما أن قيمة | Z | المحسوبة > القيمة الحرجة :

| ١,٤٤٢ | > ٢,٥٧٥

أي أن قيمة Z المحسوبة تقع في منطقة عدم رفض H_0 ، فإننا لا نرفض فرض العدم بمستوى معنوية ٠,٠١ . أي أننا لا نرفض الفرض القائل بأن نسبة المدخنين بين طلبة الجامعة هي ٠,٣٨ .



شكل (١٣)

تمارين (٣)

١ ـ تعاقد أحد مزارع الدواجن على توريد شحنة دجاج لأحد المطاعم ويدعي صاحب هذه المزرعة أن متوسط وزن الدجاجة لا يقل عن ١,٢٥ كيلو جرام بانحراف معياري ٠,١٠ كيلو جرام . ولقد قام صاحب المطعم بسحب عينة عشوائية من ١٠٠ دجاجة فوجد أن متوسط وزنها ١,١٠ كيلو جرام . والمطلوب :

١ ـ هل يجب على صاحب المطعم رفض الشحنة ؟ حل باستخدام طريقة القيمة الحرجة ثم باستخدام طريقة القيمة الاحتمالية ، استخدم مستوى معنوية ٠,٠٥ .

٢ ـ أوجد الخطأ من النوع الثاني واحسب قوة الاختبار إذا علمت أن فرض العدم غير صحيح وأن متوسط وزن الزجاجة الحقيقي هو ١,٢٤ كيلو جرام .

٢ ـ في أحد مصانع الإطارات كان العمر الافتراضي لها يتبع التوزيع المعتدل بمتوسط ٣١٠٠٠ كيلو متر . ولقد قامت إدارة البحوث بالمصنع بإضافة مادة جديدة لزيادة العمر الافتراضي لهذه الإطارات . وللتأكد من ذلك تم تصنيع ١٥ إطاراً بإضافة هذه المادة الجديدة ، وبقياس أعمار هذه الإطارات تبين أن الوسط الحسابي لعمر الإطار هو ٣٣٠٠٠ كيلو متر بانحراف معياري ٢٠٥٠ كيلو متر . والمطلوب : هل يجب تعميم إضافة هذه المادة الجديدة على جميع إطارات المصنع ؟ استخدم مستوى معنوية ٠,٠١ .

٣ ـ تقوم أحد الآلات بتعبئة المياه المعدنية في زجاجات على أن تحتوي الزجاجة على ١,٥ كيلو جرام من المياه المعدنية ولقد قام مراقب جودة

الإنتاج بسحب عينة من ٢٠ زجاجة فوجد أن الوسط الحسابي والانحراف المعياري لوزن المياه المعدنية في الزجاجة كانا على التوالي : ١,٣٧ كيلو جرام ، ٠,٢٠ كيلو جرام . ولقد استنتج مراقب الجودة أن وزن العبوة مختلف عن المواصفات ، وبافتراض أن المجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً ، فالمطلوب : هل توافق مراقب الجودة في الرأي ؟ استخدم مستوى معنوية ٠,٠٥.

٤ ـ في أحد المصانع تقوم آلة بإنتاج مواسير من الألومنيوم أقطارها ٣ سم ولقد قام مراقب جودة الإنتاج بسحب عينة عشوائية من ٩ مواسير فوجد أقطارها كالآتي :

٢,٧ ٣,٢ ٣,٤ ٢,٩ ٣,١ ٢,٨ ٣,٣ ٣,٦ ٣,٧

فإذا علمت أن أقطار المواسير يتبع التوزيع الطبيعي تقريباً ، فالمطلوب : باستخدام مستوى معنوية ٠,٠١ تحديد ما إذا كانت هذه الآلة تحتاج إلى صيانة أم لا ؟

٥ ـ تعلن إحدى الشركات المنتجة لبطاريات السيارات أن العمر الافتراضي للبطارية هو ٤٠ شهر . ولقد قام أحد المستوردين لهذه البطاريات بسحب عينة عشوائية من ٢٠ بطارية فوجد أن متوسط عمر البطارية ٣٧,٨ شهر بانحراف معياري ٣,٥ شهر . وبافتراض أن المجتمع الذي سحبت منه العينة يتوزع توزيعاً معتدلاً . فالمطلوب : هل توافق على أن عمر هذه البطاريات أقل من ٤٠ شهر ؟ استخدم مستوى معنوية ٠,٠٥.

٦ ـ قامت إحدى شركات السجائر بإنتاج نوع جديد وأعلنت أن القطران لا يزيد عن ٣,٨ ملليجرام في السيجارة . ولقد قام أحد المستوردين بسحب عينة عشوائية من ٢٠ سيجارة فوجد أن الوسط الحسابي لكمية القطران بها هو ٤,١٤ ملليجرام بانحراف معياري ع = ٠,٤ ملليجرام .

ولقد قام المستورد برفض الشحنة . وبافتراض أن المجتمع الذي سحبت منه هذه العينة يتوزع توزيعاً معتدلاً ، المطلوب : هل تتفق مع المستورد في الرأي ؟ استخدم مستوى معنوية ٠,٠١ .

٧ ــ تقوم إحدى الشركات بإرسال البريد السريع ، وتعلن هذه الشركة بأنها توصل على الأقل ٦٥ ٪ من البريد في ظرف ٤٨ ساعة . وتقوم إدارة الرقابة على الجودة بالتأكد من وقت إلى أخر بأن البريد يصل في ميعاده . ولقد أخذت هذه الإدارة عينة من ٢٠٠ طرد بريدي ووجدت أن ١٢٠ منها وصلت في خلال ٤٨ ساعة . والمطلوب :

أ ــ بأخذ مستوى معنوية ٠,٠١ هل تتفق مع الشركة في الرأي ؟

ب ــ ما هو استنتاجك في (أ) إذا كان احتمال الحصول على الخطأ من النوع الأول يساوي صفر ؟ فسر الإجابة .

٨ ــ تقوم إحدى شركات الكمبيوتر بإنتاج اسطوانات الكمبيوتر (ديسكات) وتقوم إحدى الآلات بتصنيع هذه الاسطوانات ، ومن المعلوم أن نسبة العادم من إنتاج هذه الآلة لا يتعدى ٥ ٪ . ولقد قام مراقب جودة الإنتاج بسحب عينة عشوائية من ٤٠٠ اسطوانة فوجد أن بها ٣٠ اسطوانة معيبة . والمطلوب :

أ ــ باستخدام مستوى معنوية ٠,٠٥ ، تحديد ما إذا كانت هذه الآلة تحتاج إلى صيانة .

ب ــ وباستخدام مستوى معنوية ٠,٠١ ، هل تستنتج نفس الاستنتاج في (أ) ؟

جـ ــ أوجد القيمة الاحتمالية p .

الفصل الرابع

أساليب الاستدلال الإحصـــائي للمقارنـة

بين معالم مجتمعين

Statistical Inference Techniques For Comparing the Parameters Of Two Populations

٤-١ مقدمة

نتناول في هذا الفصل بعض أساليب الاستدلال الإحصائي والتي تناسب ما يسمى بدراسات المقارنة، حيث يكون لدينا في هذه الحالة مجتمعين نهتم بعقد مقارنات بين معالمهما المتناظرة وذلك بهدف التعرف على أوجه الاختلاف والتشابه بينهما. فعلى سبيل المثال، قد نهتم في دراسة معينة بمقارنة أجور العمال وأجور العاملات للتعرف على ما إذا كان هناك تمييز في مستويات الأجور بين النوعين أم لا. كذلك قد نهتم بمقارنة درجات الطلاب والطالبات في امتحان معين، أو مقارنة درجات نفس المجموعة من الطلبة في امتحانين مختلفين. وقد نهتم في دراسة ثالثة بمقارنة تأثير أسلوبين مختلفين لعلاج مرض معين على الحالة الصحية للمصابين بهذا المرض، أو مقارنة الحالة الصحية لمجموعة من المرضى

قبل وبعد تناول دواء معين. ويمكن القول بأن دراسات المقارنة هي الأكثر انتشارا في الحياة العملية نظرا لتشعب مجالات تطبيقها.

عند إجراء دراسة تتضمن المقارنة بين مجتمعين، فلابد في البداية من تحديد العلاقة بين المجتمعين. ونميز في هذا الصدد بين حالتين. الحالة الأولى هي حالة استقلال مشاهدات كل مجتمع عن مشاهدات المجتمع الآخر. والحالة الثانية هي حالة ارتباط كل مشاهدة في المجتمع الأول بمشاهدة مقابلة لها في المجتمع الثاني.

حالة استقلال مجتمعي الدراسة:

سوف نعتبر دائما أن مجتمعي الدراسة مستقلان عن بعضهما البعض إذا كانت مشاهدات أحد المجتمعين لا تتأثر ولا تعتمد على قيم ومشاهدات المجتمع الثاني. فعلى سبيل المثال، عند الحديث عن توزيع الدرجات في امتحان معين، يخضع فيه الطلبة لمراقبة جيدة، تعبر درجة كل طالب عن مستواه العلمي ومدى الجهد الذي بذله في التحضير للامتحان، وبالتالي تكون درجات الطلاب مستقلة عن بعضها البعض ومستقلة أيضا عن درجات الطالبات. كذلك عند إعطاء دواء "A" لمريض بمرض معين، وإعطاء دواء "B" لمريض آخر بنفس المرض، فإن استجابة المريض الأول للدواء "A" لا تتأثر ولا تعتمد على استجابة المريض الثاني للدواء "B". وبالتالي فإن مجتمع بيانات الحالة الصحية للمرضى الذين تتم معالجتهم بالدواء "A" يكون مستقلا عن مجتمع بيانات الحالة الصحية للمرضى الذين تتم معالجتهم بالدواء "B".

حالة عدم استقلال مجتمعي الدراسة

إذا وجدنا في أية دراسة أن كل مشاهدة في أحد المجتمعين ترتبط بمشاهدة مناظرة في المجتمع الثاني، فإننا نقول بأن مجتمعي الدراسة غير مستقلين ويوجد بينهما ارتباط. وينشأ الارتباط بين مجتمعين للبيانات بأحد أسلوبين أساسيين هما أسلوب القياس على نفس مفردات الدراسة قبل وبعد إخضاع المفردات لمعالجة معينة. وفي الأسلوب الآخر يتم تكوين أزواج متماثلة من المفردات لها تقريبا نفس السمات والخصائص حيث يتم اختيار أحد المفردات من كل زوج بطريقة عشوائية ويتم إخضاعه لمعالجة معينة بينما تخضع المفردة الثانية في كل زوج لمعالجة أخرى حيث يكون الهدف في النهاية هو عقد مقارنة بين تأثيري المعالجتين على المفردات. ووجود التماثل والتشابه بين مفردتي كل زوج يخلق نوعا من الارتباط بين المشاهدات وبالتالي نحكم على هذه الحالة بعدم استقلال مجتمعي الدراسة.

وللتوضيح نعطي فيما يلي أمثلة تطبيقية توضح مفهوم عدم الاستقلال.

المشاهدات قبل وبعد المعالجة:

افترض أننا نريد تقييم نظام غذائي معين لإنقاص الوزن. في هذه الحالة نعتبر أن أفضل أسلوب لإجراء هذه الدراسة هو أن نقوم باختيار مجموعة من الأفراد ممن يرغبون في اتباع هذا النظام ونسجل أوزانهم قبل بدء التجربة، ثم نترك كل فرد منهم يطبق هذا النظام لفترة زمنية محددة. ونقوم في نهاية التجربة بقياس الأوزان مرة أخرى ونأخذ حجم التغير في الوزن كمؤشر يعكس مدى فاعلية النظام الغذائي المقترح على تخفيض الوزن. في هذه الحالة نقول بأن بيانات أوزان جميع الأفراد الذين يمكن أن يطبقوا هذا النظام تمثل مجتمع، بينما تمثل بيانات أوزانهم بعد تطبيق النظام الغذائي

المجتمع الآخر. ولبيان أن مشاهدات المجتمعين مرتبطة ببعضها، افترض أننا بدأنا بفرد كان وزنه قبل تطبيق النظام هو ١٢٠كجم. مهما كانت فاعلية النظام الغذائي نتوقع أن يظل وزنه بعد النظام مرتفعا أيضا وليكن ١٠٠كجم على سبيل المثال. من جهة أخرى، إذا بدأنا بشخص وزنه ٨٠ كجم قد يصبح وزنه، على سبيل المثال، بعد تطبيق النظام ٧٥ كجم. أي أننا في مثل هذا النوع من التجارب نتوقع أن ترتبط القيم المرتفعة ببعضها البعض وأن ترتبط القيم المنخفضة ببعضها البعض. وبالتالي لا يوجد استقلال بين بيانات المجتمعين نتيجة أخذ القياسات على نفس الأفراد قبل وبعد التجربة.

أزواج المشاهدات المتماثلة

افترض أننا نريد المقارنة بين أسلوبين مختلفين لتدريس مقرر معين على مستوى استيعاب الطلبة ودرجاتهم في امتحان لهذا المقرر. يقتضي إجراء هذه الدراسة أن يتم تدريس المقرر لمجموعة من الطلبة باستخدام الأسلوب الأول ولمجموعة أخرى باستخدام الأسلوب الثاني (لماذا لا نستخدم نفس المجموعة من الطلبة كما في الحالة السابقة؟). إذا قلنا بأننا سوف نقوم باختيار عينة عشوائية من الطلبة لكل أسلوب من أسلوبي التدريس وفي النهاية نعقد لهم امتحان موحد، قد نصل إلى نتائج مضللة. فعلى سبيل المثال، قد نجد أن درجات الطلبة الذين درسوا المقرر بالأسلوب الأول أفضل من تلك الخاصة بالذين درسوا المقرر بالأسلوب الثاني، ولكن هذا الاختلاف قد يكون راجعا إلى تفوق طلبة المجموعة الأولى على طلبة المجموعة الثانية، وليس نتيجة تأثير أسلوب التدريس الأول. وإذا قلنا أنه يجب أن يكون مستوى جميع الطلبة في المجموعتين متقارب، نجد أنه من الصعوبة

في العديد من الأحيان الحصول على عدد كبير من الطلبة ذوي المستويات المتقاربة لتقسيمهم عشوائيا على المجموعتين. في مثل هذه الحالات نلجأ، كما سبق أن ذكرنا، إلى تكوين أزواج من المفردات المتجانسة، فنقوم بتكوين مجموعة من الأزواج، بحيث يشتمل كل زوج على طالبين في نفس المرحلة الدراسية ولهما تقريبا نفس المستوى العلمي ونفس درجة الالتزام في حضور المحاضرات وما إلى غير ذلك من العوامل. بعد ذلك، نقوم باختيار أحد الطلبة من كل زوج بطريقة عشوائية ونرسله ليدرس المقرر بالأسلوب الأول ونرسل الآخر ليدرس المقرر بالأسلوب الثاني. وفي نهاية التجربة نعقد امتحان موحد لجميع الطلبة. ولبيان كيف أنه يوجد ارتباط بين مجموعتي البيانات، نقول بأنه إذا كان لدينا زوج المستوى العلمي للطالبين فيه مرتفع، فإننا نتوقع أن يحصل كل منهما على درجات مرتفعة في الامتحان وذلك بصرف النظر عن طريقة التدريس. كذلك إذا كان لدينا زوج المستوى العلمي للطالبين فيه منخفض، نتوقع أن يحصل كل منهما على درجات منخفضة وذلك بصرف النظر عن أسلوب التدريس. أي أننا مرة أخرى نتوقع أن ترتبط المشاهدات المرتفعة ببعضها البعض والمنخفضة ببعضها البعض ويكون لدينا حالة عدم استقلال بين المجتمعين.

وترجع أهمية التمييز بين مجتمعي الدراسة من حيث الاستقلال أو عدم الاستقلال إلى أن أساليب الاستدلال الإحصائي المستخدمة تختلف في الحالتين. وسوف تشمل دراستنا في هذا الفصل المقارنة بين متوسطي مجتمعين (مستقلين أو غير مستقلين) والمقارنة بين نسبتي مجتمعين مستقلين وكذلك المقارنة بين تبايني مجتمعين مستقلين. وكما فعلنا في

الفصول السابقة، سوف نبدأ دائما بتحديد مقدرات النقاط، ثم توزيع المعاينة لهذه المقدرات ثم نتناول أساليب اختبارات الفروض وتقدير فترات الثقة.

٤-٢ الاستدلال عن متوسطي مجتمعين مستقلين

نعلم من دراستنا السابقة أن مقاييس النزعة المركزية تعكس المستوى العام لقيم الظواهر التي تتم دراستها. فإذا أردنا عقد مقارنة بين مستوى القيم في مجتمعين مستقلين، فيمكن أن يتم ذلك بمقارنة متوسطي المجتمعين لتحديد ما إذا كانا متساويين، أم أن مستوى القيم في أحد المجتمعين أعلى منه في المجتمع الآخر. وعندما يتعذر استخدام أسلوب الحصر الشامل، تكون قيمة متوسط كل من المجتمعين مجهولة وبالتالي يتم تقديرها باستخدام متوسط عينة عشوائية مسحوبة من كل مجتمع وبعد ذلك نقوم بتطبيق مجموعة من أساليب الاستدلال الإحصائي للمقارنة بين المجتمعين مماثلة لما سبق وقدمناه في الفصول السابقة.
وفي حالة دراسة مجتمعين أو أكثر تظهر الحاجة إلى استخدام أدلة سفلية تميز مقاييس المجتمع الأول عن مقاييس المجتمع الثاني كما يتضح من الجدول التالي.

المقياس	المجتمع الأول	المجتمع الثاني
متوسط المجتمع	μ_1	μ_2
الفرق بين متوسطي المجتمعين	$\Delta = \mu_1 - \mu_2$ أو $\Delta = \mu_2 - \mu_1$	
تباين المجتمع	σ_1^2	σ_2^2
التباين المشترك	σ^2	
متوسط العينة	$\bar{س}_1$	$\bar{س}_2$
حجم العينة	$ن_1$	$ن_2$
تباين العينة	$ع_1^2$	$ع_2^2$
التباين المشترك	$ع_م^2 = \frac{(ن_1-1)ع_1^2 + (ن_2-1)ع_2^2}{ن_1+ن_2-2}$	

٤-٢-١ مقدر نقطة للفرق بين متوسطي المجتمعين

استخدمنا من قبل متوسط العينة كمقدر نقطة لمتوسط المجتمع. وبالتالي يكون $\bar{س}_1$ مقدر نقطة لمتوسط المجتمع الأول μ_1، ويكون $\bar{س}_2$ مقدر نقطة لمتوسط المجتمع الثاني. ونستخدم في هذا الفصل الفرق بين متوسطي العينتين كمقدر نقطة للفرق بين متوسطي المجتمعين.

مقدر نقطة للفرق $\Delta = \mu_1 - \mu_2$، يكون الفرق بين متوسط العينة الأولى ومتوسط العينة الثانية ($\bar{س}_1 - \bar{س}_2$) . ومن ناحية أخرى، يكون

مقدر نقطة للفرق $\Delta = \mu_2 - \mu_1$، هو الفرق بين متوسط العينة الثانية والأولى ($\overline{س}_2 - \overline{س}_1$). وحيث أن كلا من $\overline{س}_1$ و $\overline{س}_2$ يكون متغيرا عشوائيا، فإن الفرق بينهما يكون هو الآخر متغيرا عشوائيا، ويكون له توزيع احتمالي هو توزيع المعاينة للفرق بين متوسطي عينتين.

٤-٢-٢ توزيع المعاينة للفرق $\overline{س}_1 - \overline{س}_2$

إذا فرضنا أن مجتمعي الدراسة مستقلان ولكل منهما توزيع طبيعي، ع(μ_1، σ_1^2) للمجتمع الأول و ع(μ_2، σ_2^2) للمجتمع الثاني، فمن نظرية 1-1 بالفصل الأول، يكون لمتوسط عينة المجتمع الأول، $\overline{س}_1$، توزيع طبيعي ع(μ_1 ، $\frac{\sigma_1^2}{ن_1}$). كذلك، يكون لمتوسط عينة المجتمع الثاني، $\overline{س}_2$، توزيع طبيعي ع(μ_2 ، $\frac{\sigma_2^2}{ن_2}$). وحيث أن مجتمعي الدراسة مستقلان، فإن المتغيرين العشوائيين يكونان مستقلان أيضا. وإذا استخدمنا الرمز $\hat{\Delta}$ للتعبير عن مقدر النقطة للفرق بين متوسط المجتمع الأول والثاني، أي أن:

$$\hat{\Delta} = \overline{س}_1 - \overline{س}_2$$

فإن $\hat{\Delta}$ يكون دالة خطية في متغيرين مستقلين ولكل منهما توزيع طبيعي، وبالتالي يكون للمتغير العشوائي $\hat{\Delta}$ توزيع طبيعي أيضا. ولتحديد متوسط وتباين هذا التوزيع لاحظ أن:

$أ_1 = 1$ و $أ_2 = -1$.

بالتالي يكون الوسط الحسابي لتوزيع $\hat{\Delta} = \overline{س}_١ - \overline{س}_٢$ من نظرية 1-1 هو:

$$\mu_{\overline{س}_١ - \overline{س}_٢} = (١)\,\mu_١ + (-١)\,\mu_٢$$

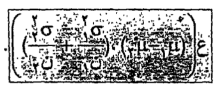
$$= \mu_١ - \mu_٢$$

ويكون تباين توزيع $\hat{\Delta}$ هو:

$$\sigma^٢_{\overline{س}_١ - \overline{س}_٢} = (١)^٢ \frac{\sigma_١^٢}{ن_١} + (-١)^٢ \frac{\sigma_٢^٢}{ن_٢}$$

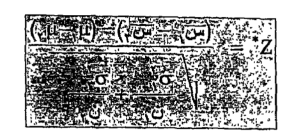
$$= \frac{\sigma_١^٢}{ن_١} + \frac{\sigma_٢^٢}{ن_٢}.$$

لاحظ أن تباين الفرق بين متوسطي العينتين هو مجموع تباينيهما. وذلك لأن التباين صيغة تربيعية تتحول فيها الإشارة السالبة إلى إشارة موجبة.
ونخلص من هذا أنه عندما يكون مجتمعا الدراسة مستقلين ولكل منهما توزيع طبيعي، يكون للفرق بين متوسطي العينتين المسحوبتين منهما توزيع طبيعي



$$ع\left((\mu_١ - \mu_٢), \left(\frac{\sigma_١^٢}{ن_١} + \frac{\sigma_٢^٢}{ن_٢}\right)\right).$$

ولحساب احتمالات أية صيغ للفرق بين متوسطي العينتين، نستخدم الوحدات المعيارية:



حيث يكون للمتغير Z^* توزيع طبيعي معياري ع(صفر ، 1).

مثال ٢-١

سحبت عينة عشوائية حجمها ٢٥ مفردة من مجتمع له توزيع طبيعي ع(٣٥ ، ١٦)، وسحبت عينة عشوائية حجمها ٥٠ مفردات من مجتمع طبيعي آخر ع(٣٥، ٨,٥) ومستقل عن المجتمع الأول. ما هو احتمال ألا يتجاوز الفرق بين متوسطي العينتين وحدتان فقط لاغير؟

الحل:

إذا فرضنا أن متوسط عينة المجتمع الأول هو $\overline{س}_١$، وأن متوسط عينة المجتمع الثاني هو $\overline{س}_٢$، يكون المطلوب هو حساب احتمال الصيغة ($|\overline{س}_١ - \overline{س}_٢| \leq ٢$). وهنا أخذنا الفرق المطلق بين متوسطي العينتين لأن المطلوب لم يتضمن أي اتجاه للفرق بينهما.

$$ح(|\overline{س}_١ - \overline{س}_٢| \leq ٢)$$

$$= ح(-٢ \leq \overline{س}_١ - \overline{س}_٢ \leq ٢)$$

من المعطيات نجد أن:

$\mu_1 - \mu_2 = ٣٧ - ٣٧ =$ صفر.

$$\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2} = \frac{١٦}{٢٥} + \frac{٨,٥}{٥٠}$$

$= ٠,٨١$.

وبتحويل حدود المتباينة في صيغة الاحتمال السابق إلى وحدات معيارية تصبح على الصورة:

$$ح(-2 \leq \bar{س}_1 - \bar{س}_2 \leq 2)$$

$$= ح\left(\frac{-٢ - صفر}{\sqrt{٠,٨١}} \leq \frac{(\bar{س}_1 - \bar{س}_2) - صفر}{\sqrt{٠,٨١}} \leq \frac{٢ - صفر}{\sqrt{٠,٨١}}\right)$$

$$= ح(-٢,٢٢ \leq Z^* \leq ٢,٢٢)$$

$$= \Phi(٢,٢٢) - \Phi(-٢,٢٢)$$

$$= ٢\Phi(٢,٢٢) - 1$$

$$= ٢ \times ٠,٩٨٦٨ - 1$$

$$= ٠,٩٧٣٦ .$$

نخلص من هذا المثال إلى أنه إذا كان لمجتمعي الدراسة نفس المتوسط، فإننا نتوقع أن يكون الفرق بين متوسطي العينتين طفيفا باحتمالات كبيرة. وهذا يتضمن أن الفرق بين متوسطي العينتين يعكس بصورة جيدة حقيقة الفرق بين متوسطي المجتمعين.

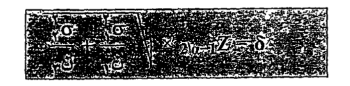
وجدنا في الفصل الاول أن حد خطأ التقدير باحتمال ١ − α لتقدير متوسط المجتمع μ يكتب على الصورة

$$\delta = Z_{1-\frac{\alpha}{2}} \times \text{الخطأ المعياري للمقدر}$$

وإذا كتبنا خطأ تقدير الفرق بين متوسطي مجتمعين باستخدام الفرق بين متوسطي عينتين مسحوبتين من هذين المجتمعين على الصورة:

$$\text{خطأ التقدير} = \left| (\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2) \right|،$$

فيمكن استخدام علاقة مماثلة لإيجاد حد خطأ تقدير الفرق بين متوسطي المجتمعين باستخدام الفرق بين متوسطي عينتين مسحوبتين من المجتمعين باحتمال $1-\alpha$. وتكون صيغة خطأ التقدير في هذه الحالة، ومن توزيع المعاينة للفرق بين متوسطي العينتين، على الصورة:



$$\delta = Z_{1-\alpha/2} \times \sqrt{\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}}$$

مثال ٤-٢

تم اختيار عينة عشوائية حجمها ٢٠ مفردة من مجتمع له توزيع طبيعي ع(μ_1 , ١٦)، وعينة عشوائية حجمها ٢٥ مفردة من مجتمع له توزيع طبيعي ع(μ_2 ، ١٠,٢٥).

المطلوب:

1- ما هو حد خطأ تقدير الفرق بين متوسطي المجتمعين باستخدام الفرق بين متوسطي العينتين عند احتمال ٩٥% ؟

2- إذا أردنا اختيار عينتين من المجتمعين لهما نفس الحجم، ن، ما هي قيمة ن التي تجعل أقصى حجم لخطأ تقدير الفرق بين متوسطي المجتمعين وحدة واحدة ونصف باحتمال ٩٩%.

الحل:

1- حد خطأ التقدير باحتمال 95%

العينة الأولى : $ن_1 = ٢٠$ $\sigma_1^2 = ١٦$

العينة الثانية : $ن_2 = ٢٥$ $\sigma_2^2 = ١٠,٢٥$

$1- \alpha = ٠,٩٥$ ، $\alpha = ٠,٠٥$ $1-\alpha\backslash٢ = ٠,٩٧٥$

$Z_{1-\alpha\backslash 2} = ١,٩٦$.

$$\delta = ١,٩٦ \times \sqrt{\frac{١٦}{٢٠} + \frac{١٠,٢٥}{٢٥}}$$

$$\delta = ١,٩٦ \times \sqrt{١,٢١}$$

$$\delta = ٢,١٥٦$$

2- تحديد حجم العينتين المتساوي

$$\delta = ١,٥ \qquad 1-\alpha = ٠,٩٩ \qquad Z_{1-\alpha\backslash 2} = Z_{0.995} = ٢,٥٧٦$$

$$١,٥ = ٢,٥٧٦ \times \sqrt{\frac{١٦}{ن} + \frac{١٠,٢٥}{ن}}$$

$$١,٥ = ٢,٥٧٦ \times \sqrt{\frac{٢٦,٢٥}{ن}}$$

إذن

$$\sqrt{ن} = \frac{٢,٥٧٦ \times ٥,١٢٣}{١,٥} = ٨,٧٩٨$$

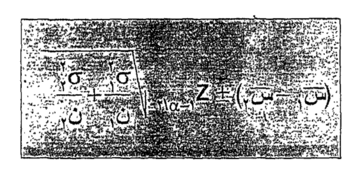
وبالتالي يكون حجم العينة التي يجب اختيارها من كل مجتمع هو:

ن = ٧٨ مفردة تقريبا.

٤-٢-٣ تقدير فترات الثقة واختبارات الفروض للفرق بين متوسطي مجتمعين طبيعيين معلومي التباين

لإيجاد مقدر فترة ثقة (١-α)% للفرق بين متوسطي مجتمعين لهما توزيعان طبيعيان تباينهما معلوم، نستخدم نفس التعريف الذي قدمناه عند تقدير فترة ثقة لمتوسط مجتمع، حيث ذكرنا أن حدود فترة الثقة تحسب بطرح وإضافة حد خطأ التقدير عند احتمال ١ - α إلى مقدر النقطة.

وبالتالي تكتب حدود فترة ثقة $(1-\alpha)\%$ للفرق بين متوسطي المجتمعين الأول والثاني $\mu_1 - \mu_2$ على الصورة:



$$(\bar{س}_1 - \bar{س}_2) \pm Z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}}$$

وإذا أردنا إيجاد مقدر فترة ثقة للفرق بين متوسطي المجتمعين الثاني والأول $\mu_2 - \mu_1$، فإننا نعكس اتجاه الطرح في صيغة فترة الثقة السابقة ليكون $(\bar{س}_2 - \bar{س}_1)$.

لإجراء اختبارات فروض تتعلق بالفرق بين متوسطي مجتمعين، فإننا نتناول المراحل الأربع السابق تقديمها في الفصل الثالث، ولكن في حالتنا الراهنة تكون معلمة الاختبار هي $\Delta = \mu_1 - \mu_2$ $(\Delta = \mu_2 - \mu_1)$. ويمكن إجراء الاختبارات لأية قيمة نظرية Δ_0، ولكننا سوف نقتصر على الحالة الأكثر شيوعا في الحياة العملية والتي تتضمن وجود أو عدم وجود فرق بين متوسطي المجتمعين. أي أننا سوف نقتصر على دراسة الحالة $\Delta_0 =$ صفر.

المرحلة الأولى: الفروض الإحصائية

في مشاكل اختبار معنوية الفرق بين متوسطي مجتمعين، تكون الفروض الإحصائية على أحد الأشكال الثلاثة السابق دراستها في اختبارات الطرفين، اختبار الطرف الأيمن واختبار الطرف الأيسر.

اختبار الطرفين:

يستخدم اختبار الطرفين في الحالة التي نريد فيها اختبار ما إذا كان للمجتمعين نفس المتوسط أم يوجد بينهما اختلاف، وذلك بصرف النظر عن اتجاه هذا الاختلاف. فعلى سبيل المثال، إذا أردنا اختبار الفرض بأنه لكل من الطلاب والطالبات نفس مستوى الأداء في امتحان مقرر معين في مقابل الفرض بأن إحدى المجموعتين أفضل من الأخرى، دون تحديد اتجاه الأفضلية، فإننا نستخدم اختبار الطرفين. ويمكن أن تكتب الفروض الإحصائية في هذه الحالة على إحدي الصور:


$H_0: \mu_1 = \mu_2$ مقابل $H_1: \mu_1 \neq \mu_2$

أو

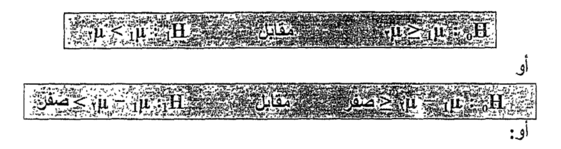
$H_0: \mu_1 - \mu_2 =$ صفر مقابل $H_1: \mu_1 - \mu_2 =$ صفر

أو:

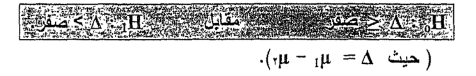
$H_0: \Delta =$ صفر مقابل $H_1: \Delta \neq$ صفر

اختبار الطرف الأيمن:

يستخدم هذا الاختبار إذا كان القرار الذي نرغب في الوصول إليه يتعلق بتحديد ما إذا كان متوسط المجتمع الأول (μ_1) يزيد بصورة جوهرية عن متوسط المجتمع الثاني (μ_2) أم لا (يساويه أو قد يقل عنه). على سبيل المثال، عند مقارنة مستوى أداء كل من الطلاب والطالبات في امتحان معين، إذا كان متوسط درجات الطلاب هو μ_1، ومتوسط درجات الطالبات هو μ_2، وأردنا اختبار ما إذا كان الطلاب أفضل من الطالبات، تكون الفروض الإحصائية على إحدى الصور:



أو:

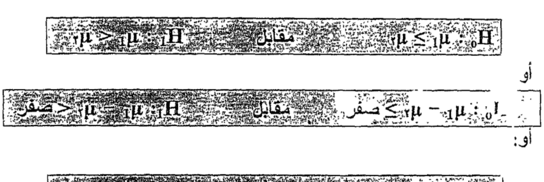


(حيث $\Delta = \mu_1 - \mu_2$).

لا حظ أن تعريف Δ يعتبر أساسي في هذه الحالة وذلك لأننا إذا ما عرفنا Δ بصورة عكسية، أي $\Delta = \mu_2 - \mu_1$ ، فإن وجهة الاختبار سوف تتغير ويصبح اختبار طرف أيسر بدلا من اختبار طرف أيمن.

اختبار الطرف الأيسر:

يستخدم هذا الاختبار إذا كان القرار الذي نرغب في الوصول إليه يتعلق بتحديد ما إذا كان متوسط المجتمع الأول (μ_1) يقل بصورة جوهرية عن متوسط المجتمع الثاني (μ_2) أم لا (يساويه أو قد يزيد عنه). على سبيل المثال، عند مقارنة مستوى أداء كل من الطلاب والطالبات في امتحان معين إذا كان متوسط درجات الطلاب هو μ_1، ومتوسط درجات الطالبات هو μ_2 ، وأردنا اختبار ما إذا كن الطالبات أفضل من الطلاب أم لا، تكون الفروض الإحصائية على إحدى الصور:



$H_0: \mu_1 \geq \mu_2$ مقابل $H_1: \mu_1 < \mu_2$

أو

$H_0: \mu_1 - \mu_2 \geq$ صفر مقابل $H_1: \mu_1 - \mu_2 <$ صفر

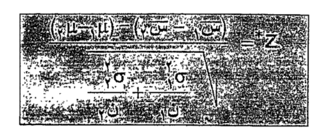
أو:

$H_0: \Delta \geq$ صفر مقابل $H_1: \Delta <$ صفر

(حيث $\Delta = \mu_1 - \mu_2$).

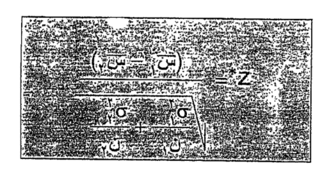
المرحلة الثانية: إحصائية الاختبار

سبق أن وجدنا من توزيع المعاينة للفرق بين متوسطي عينتين مسحوبتين من توزيعين طبيعيين ومستقلين أن المتغير العشوائي



$$Z^* = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}}}$$

يكون له توزيع طبيعي معياري. وإذا ما استخدمنا الصيغة السابقة كإحصائية للاختبار، فإننا نضع الفرق $\mu_1 - \mu_2$ مساويا للصفر وذلك لأننا وكما ذكرنا من قبل نجري اختبارات الفروض في ظل صحة فرض العدم. وبالتالي تصبح إحصائية الاختبار على الصورة:



$$Z^* = \frac{(\bar{س}_1 - \bar{س}_2)}{\sqrt{\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}}}$$

ويلاحظ أن البسط في صيغة Z^* يشتمل على الفرق بين متوسطي العينتين. وكلما قل الفرق بين المتوسطين، وبالتالي قلت القيمة العددية لإحصائية الاختبار واقتربت من الصفر، كلما كان لدينا دلالة أقوى على عدم وجود فرق بين متوسطي المجتمعين وعلى صحة فرض العدم. ومن ناحية أخرى، كلما زاد الفرق بين متوسطي العينتين وبالتالي زادت القيمة العددية لإحصائية الاختبار، كلما زادت الدلالة على عدم صحة فرض العدم.

المرحلة الثالثة: تحديد القيمة الحرجة

حيث أننا في هذه المرحلة نفترض أن تباين كل من المجتمعين معلوم، فإننا نقوم بتحديد القيمة الحرجة من جداول التوزيع الطبيعي المعياري. وهي كما سبق $Z_{1-\alpha}$ في حالة اختبار الطرف الأيمن، $-Z_{1-\alpha}$ في حالة اختبار الطرف الأيسر، و $\pm Z_{1-\alpha\backslash2}$ في حالة اختبار الطرفين.

المرحلة الرابعة: تقرير نتيجة الاختبار

إذا وقعت قيمة إحصائية الاختبار داخل المنطقة الحرجة، أي إذا زادت القيمة العددية لإحصائية الاختبار عن القيمة الجدولية، فإننا نرفض فرض العدم ونقبل الفرض البديل ونستنتج وجود اختلاف بين متوسطي المجتمعين أو أن متوسط أحدهما يكون أكبر من متوسط الآخر.

كذلك يمكن تقرير نتيجة الاختبار بحساب القيمة الاحتمالية بنفس الأسلوب السابق شرحه في الفصل السابق.

مثال ٤-٣

في عينة من ٢٠٠ عامل من العاملين بشركات المقاولات والإنشاءات وجد أن المتوسط الشهري لأجر العامل هو ٥٣٨ جنيها. وفي عينة من ١٥٠ عاملا بالشركات الصناعية وجد أن المتوسط الشهري لأجر العامل هو ٥١٢ جنيها. وعلى فرض أن توزيع أجور العمال في كل من شركات القطاعين هو توزيع طبيعي بانحراف معياري ٦٦ و ٦٠ جنيها على الترتيب، فالمطلوب:

١- ما هو مقدر نقطة للفرق بين متوسطي أجور العمال في قطاع المقاولات وقطاع الصناعة؟

٢-أوجد تقدير نقطة للفرق بين متوسطي الأجور في قطاعي المقاولات والصناعة.

٣-أوجد تقدير فترة ثقة ٩٥% للفرق بين متوسطي الأجور في المطلوب السابق.

٤-اختبار الفرض بأن مستوى أجور العمال يزيد بصورة جوهرية في قطاع المقاولات عنه في قطاع الصناعة وذلك عند مستوى معنوية ٥%.

الحل:

يلاحظ أن لكل قطاع طبيعة خاصة تحدد مستويات أجوره، وبالتالي نعتبر أن مجتمعي الدراسة مستقلين عن بعضهما البعض.

	مقاولات	صناعة
متوسط المجتمع	μ_1	μ_2
الانحراف المعياري	σ_1	σ_2
تباين المجتمع	$\sigma_1^2 = ٤٣٥٦$	$\sigma_2^2 = ٣٦٠٠$
متوسط العينة	$\overline{س}_1 = ٥٣٨$	$\overline{س}_2 = ٥١٢$
حجم العينة	$ن_1 = ٢٠٠$	$ن_2 = ١٥٠$

1- من الفصل الأول يكون مقدر النقطة هو صيغة إحصاء العينة

المستخدمة لتقدير المعلمة. وبالتالي يكون مقدر نقطة للفرق بين متوسطي المجتمعين هو الفرق بين متوسطي العينتين. ومن السؤال نجد أن المطلوب

هو مقدر نقطة للفرق $\mu_{١} - \mu_{٢}$

$$\widehat{\mu_1 - \mu_2} = \hat{\mu}_{١} - \hat{\mu}_{٢} = \overline{س}_{١} - \overline{س}_{٢}.$$

2- من الفصل الأول يكون تقدير النقطة هو قيمة المقدر محسوبة من بيانات العينة المتاحة. وبالتالي يكون تقدير النقطة للفرق بين متوسطي الأجور حسب المطلوب هو:

$$\widehat{\mu_1 - \mu_2} = ٥٣٨ - ٥١٢ = ٢٦ \text{ جنيها.}$$

3- تقدير فترة ثقة للفرق بين متوسطي المجتمعين

حيث أن توزيعي المجتمعين هما توزيعان طبيعيان معلوما التباين، ومجتمعي الدراسة مستقلان، فإننا نقوم بتقدير فترة الثقة وإجراء اختبار الفروض في المطلوب التالي باستخدام التوزيع الطبيعي المعياري.

$$١ - \alpha = 0.95 \quad \alpha = 0.05 \quad Z_{1-\alpha\backslash 2} = Z_{0.975} = 1.96$$

$$٢٦ \pm ١,٩٦ \sqrt{\frac{٤٣٥٦}{٢٠٠} + \frac{٣٦٠٠}{١٥٠}}$$

$$٢٦ \pm ١,٩٦ \times ٦,٧٧$$

$$٢٦ \pm ١٣,٢٦$$

الحد الأدنى للفرق بين المتوسطين = ١٢,٧٤ جنيها.

الحد الأعلى للفرق بين المتوسطين = ٣٩,٢٦ جنيها.

4- فرض البحث الذي نريد اختباره يتعلق بما إذا كان متوسط أجور العمال في قطاع المقاولات، μ_1، يزيد جوهريا عنه في قطاع الصناعة $\mu_٢$ ، أم لايزيد (يساوي أو قد يقل). وبالتالي يكون الاختبار في هذه الحالة هو اختبار طرف أيمن.

الفروض الإحصائية:

$H_0 : \mu_1 \leq \mu_٢$ مقابل $H_1 : \mu_1 > \mu_٢$

إحصائية الاختبار

$$Z^* = \frac{(\overline{س}_١ - \overline{س}_٢)}{\sqrt{\frac{\sigma_١^٢}{ن_١} + \frac{\sigma_٢^٢}{ن_٢}}}$$

$$Z^* = \frac{٥٣٨ - ٥١٢}{\sqrt{\frac{٤٣٥٦}{٢٠٠} + \frac{٣٦٠٠}{١٥٠}}}$$

$$= \frac{٢٦}{٦.٧٧} = ٣,٨٤$$

القيمة الحرجة

$\alpha = 0.05$ - اختبار طرف أيمن

$Z_{1-\alpha} = Z_{0.95} = 1.645$

نتيجة الاختبار

حيث أن قيمة إحصائية الاختبار تزيد عن القيمة الحرجة، فإننا نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن متوسط أجور العمال في قطاع المقاولات يزيد عن متوسط أجور العمال في قطاع الصناعة وذلك عند مستوى معنوية ٥%.

٤-٢-٤ تقدير فترات الثقة واختبارات الفروض للفرق بين متوسطي مجتمعين غير معلومي التباين.

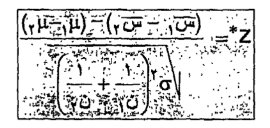
سبق أن رأينا في الفصلين الثاني والثالث عند الاستدلال عن متوسط مجتمع غير معلوم التباين أننا نقوم باستخدام تباين العينة، ع٢ كمقدر نقطة لتباين المجتمع. ويكون توزيع المعاينة لإحصائية الاختبار عندئذ هو توزيع t وليس التوزيع الطبيعي المعياري. وفي الفصل الحالي أيضا عند الاستدلال عن الفرق بين متوسطي مجتمعين لهما توزيع طبيعي، إذا كان تباين المجتمعين غير معلوم، فإننا سوف نقوم باستخدام توزيع t. وسوف نفرق في هذا الصدد بين حالتين هما، الحالة التي يكون فيها

للمجتمعين تباين متساوي، أي $\sigma_1^2 = \sigma_2^2 = \sigma^2$، والحالة التي يكون فيها

للمجتمعين تباين غير متساوي، أي $\sigma_1^2 \neq \sigma_2^2$.

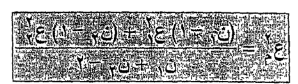
الحالة الأولى: تباين المجتمعين متساوي

إذا أمكننا افتراض أن لمجتمعي الدراسة نفس التباين، سواء كان هذا الافتراض معلوما نظريا أو تم اختباره، كما سنرى في نهاية هذا الفصل، تكتب صيغة Z^* من توزيع المعاينة للفرق بين متوسطي مجتمعين على الصورة:



$$Z^* = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2\left(\frac{1}{ن_1} + \frac{1}{ن_2}\right)}}$$

وإذا كانت قيمة σ^2 غير معلومة، فإننا نقوم بتقديرها من البيانات المتاحة. في هذه الحالة، يعتبر تباين العينة الأولى، $ع_1^2$، مقدر نقطة لتباين المجتمع σ^2. وكذلك يكون تباين العينة الثانية، $ع_2^2$، مقدر نقطة لتباين المجتمع. وحيث أن تباين المجتمعين متساوي، فإننا نقوم باستخدام متوسط مرجح لتباين العينتين معا (حيث يكون الترجيح على أساس درجات الحرية لكل عينة)، لتقدير σ^2. ويطلق على مقدر تباين المجتمعين المتساوي اسم التباين المشترك ونرمز له بالرمز $ع_م^2$ ويحسب باستخدام العلاقة:

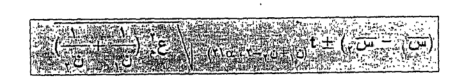


ويمكن أيضا استخدام الصيغة المكافئة التالية

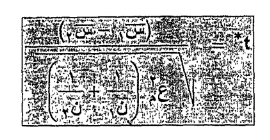


وعند استبدال σ^2 في صيغة Z السابقة بالمقدر $ع_م^2$، فإننا نحصل على متغير عشوائي جديد هو t له توزيع t بدرجات حرية ($ن_1 + ن_2 - 2$) ونستخدمه كما ذكرنا في بناء فترات الثقة وإجراء اختبارات الفروض للفرق بين متوسطي المجتمعين.

ويكون مقدر فترة ثقة $(1 - \alpha)\%$ للفرق بين متوسطي المجتمعين الاول والثاني $\mu_1 - \mu_2$ على الصورة:



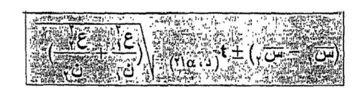
وفي اختبارات الفروض، تبقى الفروض الإحصائية كما هي على أحد الصور الثلاث السابق تقديمها، وتصبح إحصائية الاختبار على الصورة:

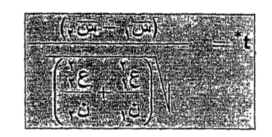


وفي المرحلة الثالثة، يتم تحديد القيمة الحرجة من جداول توزيع t بدرجات حرية ($ن_1 + ن_2 - 2$) وعند الاحتمال المناسب حسب نوع الاختبار. ويتم تقرير نتيجة الاختبار كالمعتاد بمقارنة قيمة إحصائية الاختبار بالقيمة الحرجة كما سبق بيانه.

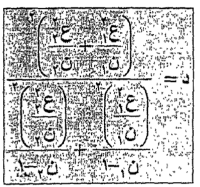
الحالة الثانية: تباين المجتمعين غير متساوي

إذا كانت هناك دلالة قوية على عدم تساوي تبايني المجتمعين، فإنه لا يصح في هذه الحالة استخدام التباين المشترك، وإنما نقوم بتقدير تباين كل مجتمع على حدة باستخدام تباين العينة الخاصة به. وتصبح صيغة كل من مقدر فترة الثقة وإحصائية الاختبار للفرق $\mu_1 - \mu_2$ على الترتيب على الصورة:





ويتمّ الكشف في جداول توزيع t عند درجات حرية د ، ومستوى المعنوية المطلوب. وتحدد درجات الحرية د بأقرب عدد صحيح أصغر من قيمة المقدار:



وفي الحالتين السابقتين عندما يكون حجم كل من العينتين كبيرا بدرجة كافية، فإننا نعود إلى استخدام التوزيع الطبيعي المعياري كتقريب لتوزيع t للعينات كبيرة الحجم.

مثال ٤-٤

للمقارنة بين مستوى كفاءة الذكور والإناث في استخدام معالج كلمات النوافذ (WINWORD)، طلب من كل فرد في عينة عشوائية من عشرة ذكور وعينة عشوائية من ثماني إناث، ممن لهم تقريبا نفس مستوى

التدريب، كتابة مقال معين باستخدام هذا البرنامج. وفي نهاية التجربة تم قياس الزمن الذي استغرقه كل منهم في الكتابة فكان متوسط الزمن المستغرق في عينة الذكور ٦,٤ دقيقة بانحراف معياري ٦٠ ثانية، وكان متوسط الزمن في عينة الإناث ٥,٢ دقيقة بانحراف معياري ٤٨ ثانية. وعلى فرض أن تباين زمن الكتابة متساوي في مجتمعي الدراسة، وأن كلا منهما يتبع توزيعا طبيعياً تقريبا، فالمطلوب:

1- هل تؤيد هذه البيانات القول بأن الإناث أفضل من الذكور في استخدام البرنامج عند مستوى معنوية 1%.
2- أوجد تقدير فترة ثقة 99% للفرق بين متوسط زمن الذكور ومتوسط زمن الإناث.

الحل:

يلاحظ في هذه المشكلة أن مجتمعي الدراسة مستقلين عن بعضهما البعض وذلك لأن الزمن الذي يستغرقه أي فرد في عينة الذكور لا تربطه أية صلة بالزمن الذي تستغرقه أية مفردة في عينة الإناث.

كذلك يجب ملاحظة أن متوسط الزمن في البيانات تم قياسه بالدقائق، بينما تم قياس الانحراف المعياري للزمن بالثواني. في هذه الحالة يجب توحيد وحدات القياس، فإما أن نحول المتوسط إلى ثواني، أو نحول الانحراف المعياري إلى دقائق.

	ذكور	إناث
متوسط الزمن في المجتمع	$\mu_{١}$	$\mu_{٢}$
تباين المجتمع (متساوي)	$\sigma^{٢}$	$\sigma^{٢}$
متوسط العينة	$\bar{\text{س}}_{١} = ٦,٤$	$\bar{\text{س}}_{٢} = ٥,٢$
الانحراف المعياري	$\text{ع}_{1} = 1$	$\text{ع}_{٢} = ٠,٨$
تباين العينة	$\text{ع}_{١}^{٢} = 1$	$\text{ع}_{٢}^{٢} = ٠,٦٤$
حجم العينة	$\text{ن}_{1} = ١٠$	$\text{ن}_{٢} = ٨$

حيث أن تباين المجتمعين متساوي، نقوم بحساب التباين المشترك

$$\text{ع}_{\text{م}}^{٢} = \frac{(١٠-١)(١)+(٨-١)(٠.٦٤)}{١٠+٨-٢} = ٠,٨٤٢٥$$

حيث أن العينتين مسحوبتان من مجتمعين طبيعيين ومستقلين وحجم كل من العينتين صغير، فإننا نستخدم توزيع t لتقدير فترة الثقة وإجراء اختبار الفروض.

1- هل الإناث أفضل من الذكور؟

تقتضي الإجابة على هذا السؤال إجراء اختبار للفروض تتم فيه صياغة فرض العدم على أساس أن الإناث ليسوا أفضل من الذكور (لا يوجد فرق في الكفاءة أو على العكس قد يكون الذكور أفضل). من جهة أخرى يتم وضع الفرض البديل على أساس أن الإناث أفضل. وحيث أن معيار الكفاءة

هنا هو استغراق زمن أقل، فإن الاختبار يكون ذو طرف أيمن إذا ما تم طرح متوسط الإناث من متوسط الذكور.

الفروض الإحصائية:

$H_0 : \mu_1 - \mu_2 \leq$ صفر $H_1 : \mu_1 - \mu_2 >$ صفر

إحصائية الاختبار

$$t^* = \frac{٦,٤ - ٥,٢}{\sqrt{٠,٨٤٢٥ \left(\frac{١}{١٠} + \frac{١}{٨}\right)}} = ٢,٧٥٦$$

القيمة الحرجة

$\alpha = 0.01$ د $= ١٠ + ٨ - ٢ = ١٦$

$t_{(0.01 ، 16)} = ٢,٥٨٣$

نتيجة الاختبار

حيث أن قيمة إجصائية الاختبار تزيد عن القيمة الحرجة، فأننا نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن الإناث أكثر كفاءة من الذكور وذلك عند مستوى معنوية 1%.

يلاحظ في هذا المثال أن قيمة إحصائية الاختبار قريبة من القيمة الحرجة. وفي الحياة العملية، يفضل في مثل هذه الحالات استخدام عينة أخرى أو زيادة حجم العينة حتى تكون لدينا دلالة كافية لصالح أحد الفرضين.

2- تقدير فترة ثقة 99% للفرق بين متوسطي زمن الذكور والإناث

$\alpha = 0.01$ $\quad \alpha \backslash 2 = 0.005$ $\quad$ د = ١٦

$t_{(0.005 ، 16)} = ٢,٩٢٤$

$$(\overline{س}_١ - \overline{س}_٢) \pm t_{(ن_١+ن_٢-٢ ، \alpha\backslash ٢)} \sqrt{ع_م^٢ \left(\frac{١}{ن_١} + \frac{١}{ن_٢}\right)}$$

$$(٦,٤ - ٥,٢) \pm ٢,٩٢٤ \times \sqrt{٠,٨٤٢٥ \left(\frac{١}{١٠} + \frac{١}{٨}\right)}$$

$$١,٢ \pm ٢,٩٢٤ \times ٠,٤٣٥٤$$

$$١,٢ \pm ١,٢٧$$

الحد الأدنى = ١,٢ – ١,٢٧ = –٠,٠٧ دقيقة.

الحد الأعلى = ١,٢ + ١,٢٧ = ٢,٤٧ دقيقة.

مثال ٤–٥

أراد مدير شركة معينة دراسة تأثير نظامين للأجور على إنتاجية العمال. يتضمن النظام الأول إعطاء أجر شهري ثابت للعمال، بينما يتضمن النظام الثاني إعطاء العامل أجر متغير يرتبط بمعدلات الإنتاج. ولإجراء هذه الدراسة، قام المدير بتطبيق النظام الأول على عينة من ٣٢ عامل يعملون في أحد فروع الشركة، فوجد أن متوسط الإنتاج الشهري للعامل هو ٢٣٥ وحدة بانحراف معياري ٨ وحدات. كذلك قام بتطبيق النظام الثاني على

عينة من ٥٠ عامل في فرع آخر من فروع الشركة، فوجد أن متوسط الإنتاج الشهري للعامل هو ٢٦٠ وحدة بانحراف معياري ٢٠ وحدة. بناء على ما تقدم، قرر المدير تطبيق نظام الأجر المتغير على كافة العاملين في جميع فروع الشركة.

1- هل تتفق معه في هذا القرار عند مستوى معنوية ٥% ؟

2- أوجد تقدير فترة ثقة 95% للفرق بين متوسطي الإنتاجية باستخدام الأجر المتغير والأجر الثابت.

الحل:

يلاحظ أن حجمي العينتين كبير في هذا المثال، وبالتالي نقوم بإجراء اختبار الفروض وتقدير فترة الثقة باستخدام توزيع Z، وذلك على الرغم من أن تباين المجتمعين مجهول. كذلك لم ينص المثال على تساوي تباينى المجتمعين، وبالتالي لا نقوم بحساب التباين المشترك وإنما نقوم بتقدير تباين كل مجتمع بتباين العينة المسحوبة منه. من جهة ثالثة، حيث أن الدراسة قد تمت في فرعين مختلفين للشركة وطبق النظام على مجموعتين مختلفتين من العمال، تكون البيانات مستقلة عن بعضها البعض في العينتين.

1- هل نظام الأجر المتغير أفضل؟

إن قرار مدير الشركة بتعميم نظام الأجر المتغير على جميع العاملين يعني أنه أفضل من حيث أنه يعطي مستوى إنتاجية أعلى. ونقوم نحن بإجراء الاختبار لتحديد ما إذا كانت الزيادة في الإنتاجية في ظل نظام الأجر المتغير

على نظيرتها في ظل نظام الأجر الثابت هي زيادة جوهرية، أم تعزى إلى عوامل الصدفة ولا يوجد اختلاف بين النظامين.

	أجر متغير	أجر ثابت
متوسط المجتمع	μ_1	μ_2
متوسط العينة	$\bar{\text{س}}_1 = ٢٦٠$	$\bar{\text{س}}_2 = ٢٣٥$
الانحراف المعياري	$\text{ع}_1 = ٢٠$	$\text{ع}_2 = ٨$
تباين العينة	$\text{ع}_1^2 = ٤٠٠$	$\text{ع}_2^2 = ٦٤$
حجم العينة	$\text{ن}_1 = ٥٠$	$\text{ن}_2 = ٣٢$

الفروض الإحصائية

يكون الاختبار وفقا لتعريفنا لبيانات الأجر المتغير على أنه يمثل بيانات المجتمع الأول، هو اختبار طرف أيمن، وذلك لأننا نضع فرض العدم على أساس أن مستوى الإنتاجية في ظل الأجر المتغير ليس أفضل (يكون مثل المتوسط في ظل الأجر الثابت وقد يكون أسوأ). من ناحية أخرى، يصاغ الفرض البديل على أن متوسط الإنتاجية في ظل نظام الأجر المتغير هو الأكبر. بالتالي تكون الفروض الإحصائية على الصورة:

$H_0 : \mu_1 \le \mu_2$ مقابل $H_1 : \mu_1 > \mu_2$

أو

$H_0 : \mu_1 - \mu_2 \le$ صفر مقابل $H_1 : \mu_1 - \mu_2 >$ صفر

إحصائية الاختبار

$$Z^* = \frac{\bar{س}_1 - \bar{س}_2}{\sqrt{\frac{ع_1^2}{ن_1} + \frac{ع_2^2}{ن_2}}}$$

$$= \frac{٢٦٠ - ٢٣٥}{\sqrt{\frac{٤٠٠}{٥٠} + \frac{٦٤}{٣٢}}} = ٧,٩$$

القيمة الحرجة

$\alpha = 0.05$ اختبار طرف أيمن $Z_{1-\alpha} = Z_{0.95} = ١,٦٤٥$

نتيجة الاختبار

حيث أن إحصائية الاختبار تزيد كثيرا عن القيمة الحرجة، نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن متوسط الإنتاجية في ظل نظام الأجر المتغير يزيد جوهريا عنه في ظل نظام الأجر الثابت وذلك عند مستوى معنوية 5%. وبالتالي نتفق مع مدير الشركة فيما توصل إليه من قرار.

القيمة الاحتمالية:

ح$(Z > ٧.٩) - 1 - \Phi(٧,٩) = 1-1 =$ صفر.

- تقدير فترة ثقة 95% للفرق بين متوسطي الإنتاجية

$$(\bar{س}_1 - \bar{س}_2) \pm 1.96 \sqrt{\frac{ع_1^2}{ن_1} + \frac{ع_2^2}{ن_2}}$$

$$٢٥ \pm ١,٩٦ \times \sqrt{١٠}$$

$$٢٥ \pm ٦,٢$$

الحد الأدنى لفترة الثقة = ٢٥ − ٦,٢ = ١٨,٨

الحد الأعلى لفترة الثقة = ٢٥ + ٦,٢ = ٣١,٢.

٤-٣ الاستدلال عن متوسطي مجتمعين غير مستقلين

تناولنا في مقدمة هذا الفصل حالتين ينشأ عنهما ارتباط وعدم استقلال مجتمعي الدراسة. في الحالة الأولى يكون المجتمع الأول هو القيم التي يمكن مشاهدتها على المفردات قبل تطبيق معالجة معينة، بينما يكون المجتمع الثاني هو القيم التي يمكن مشاهدتها على نفس المفردات بعد تطبيق المعالجة. وفي الحالة الثانية، يتم تقسيم مفردات مجتمع الدراسة إلى أزواج تتماثل وتتشابه في خصائصها حيث تخضع إحدى المفردتين في كل زوج عشوائيا لمعالجة معينة بينما تخضع المفردة الثانية لمعالجة أخرى.

كما في حالة الاستقلال يكون مقدر نقطة لمعلمة الفرق بين متوسطي المجتمعين، $\Delta = \mu_1 - \mu_2$ ، هو الفرق بين متوسطي العينتين، أو متوسط الفرق بين المشاهدات المتناظرة في العينتين. وإذا رمزنا

لمشاهدات العينة الأولى بالرمز س ولمشاهدات العينة الثانية بالرمز ص، تحسب فروق المشاهدات ونرمز لها بالرمز ف.

س	ص	ف = س - ص
$س_1$	$ص_1$	$ف_1 = س_1 - ص_1$
$س_2$	$ص_2$	$ف_2 = س_2 - ص_2$
...	....	....
...	....	....
$س_ن$	$ص_ن$	$ف_ن = س_ن - ص_ن$

ويتم إجراء اختبارات الفروض وتقدير فترات الثقة باستخدام الفروق السابقة، ونقوم بتطبيق أساليب الاستدلال المتعلقة بمتوسط مجتمع واحد (مجتمع الفروق بين مشاهدات المجموعتين) والتي سبق أن تناولناها في الفصلين الثاني والثالث.

مقدر النقطة

$$\hat{\Delta} = \bar{ف} = \frac{مجـ ف}{ن} = \bar{س} - \bar{ص}$$

وإذا كان لكل من مجتمعي الدراسة توزيع طبيعي، يكون لمجتمع الفروق توزيع طبيعي أيضا متوسطه هو Δ ونرمز لتباينه بالرمز $\sigma^2_{ف}$. وهناك علاقة بين تباين مجتمع الفروق، $\sigma^2_{ف}$، وتبايني المجتمعين σ^2_1 و σ^2_2 بالإضافة إلى معلمة أخرى يطلق عليها اسم التغاير بين متغيري الدراسة،

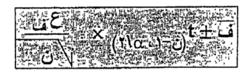
ولكننا لن نستخدم هذه العلاقة لإيجاد مقدر نقطة لمعلمة تباين الفرق نظرا لعدم دراسة التغاير بعد، ولكننا سوف نقوم هنا بتقدير تباين الفرق باستخدام تباين الفروق بين مشاهدات العينتين.

$$\hat{\sigma}^{2}_{ف} = ع^{2}_{ف} = \frac{1}{ن - 1} \text{ مجـــ } (ف - \bar{ف})^{2}$$

أو

$$ع^{2}_{ف} = \frac{1}{ن - 1}\left[\text{مجـــ } ف^{2} - \frac{(\text{مجـــ } ف)^{2}}{ن}\right]$$

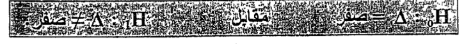
وعندما يكون حجم العينتين (ودائما ما يكون متساوي نظرا لارتباط المشاهدات في هذه الحالة) صغيرا، فإننا نستخدم توزيع t لتقدير فترات الثقة واختبارات الفروض كما سبق. ويكون مقدر فترة ثقة $(1-\alpha)\%$ للفرق بين متوسطي المجتمعين الأول والثاني



$$\bar{ف} \pm t_{(ن-1، \alpha/2)} \times \frac{ع_{ف}}{\sqrt{ن}}$$

وعند إجراء اختبارات الفروض، إذا كتبنا الفروض الإحصائية بدلالة المعلمة Δ تكون الفروض على إحدى الصور

اختبار ا لطرفين:

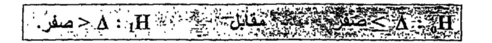


$H_0 : \Delta =$ صفر   مقابل $H_1 : \Delta \neq$ صفر

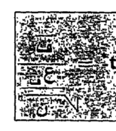
اختبار الطرف الأيمن:



اختبار الطرف الأيسر



إحصائية الاختبار



القيمة الحرجة

يتم تحديد القيمة الحرجة، في حالة العينات صغيرة الحجم، من جدول توزيع t عند درجات حرية (ن – 1) والاحتمال المناسب حسب نوع الاختبار. وفي حالة العينات كبيرة الحجم نستخدم جداول التوزيع الطبيعي المعياري.

نتيجة الاختبار

إذا زادت القيمة العددية لإحصائية الاختبار عن القيمة الجدولية، نقوم برفض فرض العدم وقبول الفرض البديل واستنتاج وجود فروق جوهرية بين بيانات مجموعتي الدراسة.

مثال ٤-٦

لقياس مدى فاعلية برنامج تدريبي معين لرفع الكفاءة الإنتاجية للعاملين، جمعت البيانات التالية عن زمن إنتاج (بالساعات) الوحدة الواحدة من منتج معين لكل عامل في عينة عشوائية قبل وبعد تنفيذ البرنامج عليهم.

رقم العامل	١	٢	٣	٤	٥	٦	٧	٨
الزمن قبل	٦	٥,٥	٥	٥	٦	٥,٨	٦,٥	٦
الزمن بعد	٥,٤	٥	٤,٦	٥	٥,٥	٥,٥	٥,٨	٦,٢

وعلى فرض أن توزيع زمن الانتاج قبل وبعد تدريب العمال هو توزيع طبيعي، فالمطلوب:

1- إيجاد تقدير فترة ثقة 95% للفرق بين متوسطي زمن إنتاج الوحدة قبل البرنامج وبعده إذا ما طبق على جميع العاملين.

2- عند مستوى المعنوية 5%، هل توصي بتطبيق هذا البرنامج التدريبي على كافة العاملين؟

الحل:

افترض أن مجتمع أزمنة الانتاج قبل البرنامج هو المجتمع الأول ومتوسطه μ_1، وأن مجتمع أزمنة الانتاج بعد البرنامج هو المجتمع الثاني ومتوسطه μ_2 وأن الفرق بين المتوسطين هو :

$$\Delta = \mu_1 - \mu_2$$

ويجب مراعاة الاتجاه السابق عند إيجاد الفروق بين مشاهدات العينتان حيث نطرح، حسب تعريف Δ، قيم زمن الانتاج بعد البرنامج من القيم المناظرة قبل البرنامج، ثم نوجد الوسط الحسابي والانحراف المعياري للفروق.

الزمن قبل	٦	٥,٥	٥	٥	٦	٥,٨	٦,٥	٦
الزمن بعد	٥,٤	٥	٤,٦	٥	٥,٥	٥,٥	٥,٨	٦,٢
ف	٠,٦	٠,٥	٠,٤	٠	٠,٥	٠,٣	٠,٧	٠,٢-
ف٢	٠,٣٦	٠,٢٥	٠,١٦	٠	٠,٢٥	٠,٠٩	٠,٤٩	٠,٠٤

$$\bar{\text{ف}} = \frac{\text{٢,٨}}{\text{٨}} = \text{٠,٣٥}$$

$$\text{ع}^{\text{٢}}_{\text{ف}} = \frac{\text{١}}{\text{ن} - \text{١}} \left[\text{مجـــ ف}^{\text{٢}} - \frac{(\text{مجـــ ف})^{\text{٢}}}{\text{ن}} \right]$$

$$= \frac{\text{١}}{\text{٧}} \left[\text{١,٦٤} - \frac{(\text{٢,٨})^{\text{٢}}}{\text{٨}} \right]$$

$$= \frac{\text{١}}{\text{٧}} \times \text{٠,٦٦} = \text{٠,٠٩٤٣}$$

$$\text{ع}_{\text{ف}} = \text{٠,٣٠٧}$$

1- تقدير فترة ثقة 95% للفرق بين متوسطي الزمن قبل وبعد تطبيق البرنامج التدريبي

$1 - \alpha = 0.95 \quad \alpha = 0.05 \quad \alpha/2 = 0.025$

د = ٨ − ١ = ٧ $\quad t_{(0.025\,،\,7)} =$ ٢,٣٦٥

$$\bar{ف} \pm t_{(ن-1\,،\,\alpha/2)} \times \frac{ع_ف}{\sqrt{ن}}$$

$$٠,٣٥ \pm ٢,٣٦٥ \times \frac{٠,٣٠٧}{\sqrt{٨}}$$

٠,٣٥ ± ٠,٢٥٧

الحد الأدنى = ٠,٣٥ - ٠,٢٥٧ = ٠,٠٩٣

الحد الأعلى = ٠,٣٥ + ٠,٢٥٧ = ٠,٦٠٧

2- لإعطاء توصية بتطبيق البرنامج أو بعدم تطبيقه، نقوم بإجراء اختبار للفروض يوضح ما إذا كان البرنامج التدريبي ذو جدوى وفائدة أم لا. وفي ضوء المعلومات المتاحة، يكون البرنامج فعال إذا كان يؤدي إلى تخفيض زمن الانتاج بصورة حقيقية وجوهرية. أي أن البرنامج يكون فعال إذا كانت القيمة الحقيقية لمعلمة الفرق Δ، وفقا لتعريفنا لها، موجبة (الفرض البديل). وعلى الجانب الآخر لا يكون البرنامج فعالا إذا بقي متوسط زمن الانتاج كما هو وأسوأ من ذلك إذا زاد زمن الانتاج بعد التدريب، وهي الحالة التي تكون فيها قيمة Δ أقل من أو تساوي الصفر (فرض العدم). وعلى ذلك يكون الاختبار في هذه الحالة هو اختبار طرف أيمن.

الفروض الإحصائية

$H_0 : \Delta \geq$ صفر مقابل $H_1 : \Delta <$ صفر.

إحصائية الاختبار

$$t^* = \frac{\bar{ف}}{\frac{ع_ف}{\sqrt{ن}}}$$

$$= \frac{٠,٣٥}{\frac{٠,٣٠٧}{\sqrt{٨}}} = ٣,٢٢٥$$

القيمة الحرجة

د = ٧ $\alpha = 0.05$ $t_{(0.05 ، 7)} =$ ١,٨٩٥

نتيجة الاختبار

حيث أن قيمة إحصائية الاختبار تزيد عن القيمة الحرجة، فإننا نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن متوسط زمن الانتاج بعد البرنامج التدريبي يقل بصورة جوهرية عما كان عليه قبل البرنامج. بناءا على ذلك تجب التوصية بتطبيق هذا البرنامج على جميع العاملين وذلك عند مستوى معنوية ٥%.

٤-٤ أساليب الاستدلال الإحصائي عن الفرق بين نسبتين في مجتمعين مستقلين

درسنا في الفصلين السابقين أساليب تقدير فترات الثقة واختبارات الفروض الإحصائية المتعلقة بنسبة ما في مجتمع إحصائي معين، وقد وجدنا أننا نعتمد على نظرية النهاية المركزية للحصول على توزيع النسبة في المجتمع. وتظهر الحاجة في العديد من الدراسات إلي إجراء اختبارات للفروض تتعلق بمعنوية الفرق بين نسبتين في مجتمعين مستقلين وإلى تقدير حدود ثقة للفرق بين النسبتين. ونتناول هذه المشكلة من مشاكل الاستدلال الإحصائي في هذا المبحث .

يتم الاستدلال عن الفرق بين نسبتين في مجتمعين مستقلين باستخدام عينتين مسحوبتين منهما. وإذا كان حجم العينة الأولى هو $ن_1$ مفردة من بينها $ك_1$ مفردة تحقق خاصية الدراسة، وكان حجم العينة الثانية هو $ن_2$ ومن بينها $ك_2$ مفردة تحقق خاصية الدراسة، فإننا نعرف بعض الرموز والمقاييس المستخدمة في عملية الاستدلال على الوجه التالي:

	المجتمع الأول	المجتمع الثاني
النسبة في المجتمع	$\theta_{١}$	$\theta_{٢}$
النسبة في العينة	$ق_{١} = \frac{ك_{١}}{ن_{١}}$	$ق_{٢} = \frac{ك_{٢}}{ن_{٢}}$

النسبة المشتركة $ق_{م} = \frac{ك_{١} + ك_{٢}}{ن_{١} + ن_{٢}}$

$$= \frac{ن_{١} \times ق_{١} + ن_{٢} \times ق_{٢}}{ن_{١} + ن_{٢}}$$

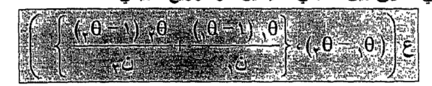
٤-٤-١ مقدر نقطة للفرق بين نسبتي المجتمعين

نستخدم الفرق بين نسبتي العينتين كمقدر نقطة للفرق بين نسبتي المجتمعين. فإذا حسبنا الفرق بين نسبتي المجتمعين بطرح نسبة المجتمع الثاني من نسبة المجتمع الأول، $\theta_{١} - \theta_{٢}$ ، يكون مقدر نقطة لمعلمة الفرق هو $ق_{١} - ق_{٢}$.

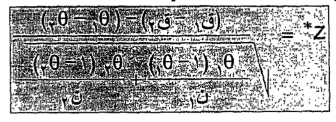
٤-٤-٢ توزيع المعاينة للفرق بين نسبتي عينتين مستقلتين

من نظرية النهاية المركزية في الفصل الأول، نعلم أن التوزيع التقريبي للنسبة $ق_{١}$ يكون توزيع طبيعي ع(θ_{1} ، $\frac{\theta_{١}(١ - \theta_{١})}{ن_{١}}$) ويكون التوزيع التقريبي للنسبة $ق_{٢}$ هو توزيع طبيعي ع($\theta_{٢}$، $\frac{\theta_{٢}(١ - \theta_{٢})}{ن_{٢}}$)

وذلك عندما يكون حجم كل عينة من العينتين كبيرا بدرجة كافية. ويمكننا من هذه المعلومات استخدام نظرية ١-٤ من الفصل الأول لنجد أن التوزيع التقريبي للفرق بين نسبتي العينتين هو توزيع طبيعي:



وبالتالي يكون للمتغير العشوائي:



توزيع طبيعي معياري ع(صفر ، 1) تقريبا.

٤-٤-٣ تقدير فترات الثقة للفرق بين نسبتين

كما نعلم يمثل المقام في الصيغة السابقة الخطأ المعياري لتوزيع المعاينة للفرق بين نسبتي العينتين. وبضربه في قيمة $Z_{1-\alpha\backslash 2}$ نحصل على حد خطأ التقدير. بالتالي نحصل مقدر فترة ثقة $(1 - \alpha)\%$ للفرق بين نسبتي المجتمعين $(\theta_١ - \theta_٢)$ بإضافة وطرح حد خطأ التقدير عند احتمال $(1-\alpha)$ إلى مقدر النقطة للفرق بين النسبتين.

$$(ق_1 - ق_2) \pm Z_{1-\alpha/2} \sqrt{\frac{\theta_1(1-\theta_1)}{ن_1} + \frac{\theta_2(1-\theta_2)}{ن_2}}$$

ويلاحظ أننا لا نستطيع حساب حدي فترة الثقة من الصيغة السابقة لتضمنها المعلمتين θ_1 و θ_2 المجهولتين، وبالتالي فإننا نستخدم مقدر النقطة لكل منهما لنحصل على صيغة مقدر فترة الثقة على الصورة:



$$(ق_1 - ق_2) \pm Z_{1-\alpha/2} \sqrt{\frac{ق_1(1-ق_1)}{ن_1} + \frac{ق_2(1-ق_2)}{ن_2}}$$

٤-٤-٤ اختبارات الفروض لمعنوية الفرق بين نسبتين

الفروض الإحصائية

إذا أردنا اختبار معنوية الفرق بين نسبتين في مجتمعين وما إذا كانت إحدى النسبتين تزيد عن الأخرى أم لا، فإن الفروض الإحصائية تكون على أحد الصور الثلاث كما سبق.

اختبار الطرفين

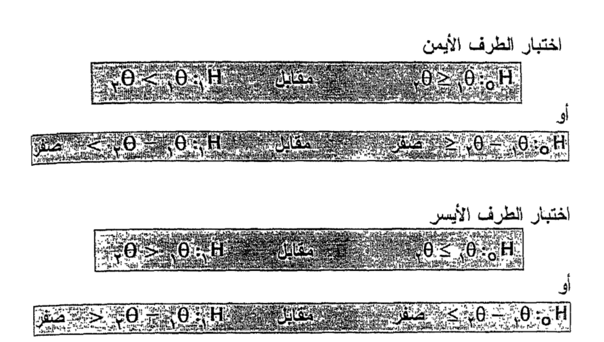


$H_0: \theta_1 = \theta_2$ مقابل $H_1: \theta_1 \neq \theta_2$

أو

$H_0: \theta_1 - \theta_2 =$ صفر مقابل $H_1: \theta_1 - \theta_2 \neq$ صفر

اختبار الطرف الأيمن



$H_0: \theta_1 \leq \theta_2$ مقابل $H_1: \theta_1 > \theta_2$

أو

$H_0: \theta_1 - \theta_2 \leq$ صفر مقابل $H_1: \theta_1 - \theta_2 >$ صفر

اختبار الطرف الأيسر

$H_0: \theta_1 \geq \theta_2$ مقابل $H_1: \theta_1 < \theta_2$

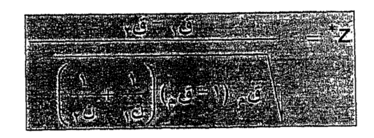
أو

$H_0: \theta_1 - \theta_2 \geq$ صفر مقابل $H_1: \theta_1 - \theta_2 <$ صفر

إحصائية الاختبار

حيث أننا نجري اختبارات الفروض دائما على أساس صحة فرض العدم، والذي يتضمن تساوي النسبتين في المجتمعين. أو أن الفرق بينهما يساوي الصفر، فإننا نقوم بوضع $(\theta_1 - \theta_2)$= صفر في صيغة Z^* السابقة.من جهة أخرى، نقوم في المقام باستخدام النسبة المشتركة لتقدير النسبتين في صيغة الخطأ المعياري على أساس تساويهما في ظل صحة فرض العدم.

من هذا تكتب صيغة إحصائية الاختبار على الصورة:



$$Z^* = \frac{\text{ق}_1 - \text{ق}_2}{\sqrt{\text{ق}_م (1 - \text{ق}_م)\left(\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2}\right)}}$$

القيمة الحرجة

تتحدد القيمة الحرجة على أساس نوع الاختبار من جداول التوزيع الطبيعي المعياري، وهي كما سبق $Z_{1-\alpha/2}$ في حالة اختبار الطرفين ، $Z_{1-\alpha}$ في حالة اختبار الطرف الأيمن و $-Z_{1-\alpha}$ في حالة اختبار الطرف الأيسر.

نتيجة الاختبار

إذا وقعت قيمة إحصائية الاختبار داخل المنطقة الحرجة، نرفض فرض العدم ونقبل الفرض البديل ونستنتج وجود اختلاف بين النسبتين في المجتمعين، أو أن أحداهما تزيد على الأخرى.

مثال ٤ –٧

تم سؤال عينة عشوائية من ٢٠٠ طالب وعينة عشوائية من ١٥٠ طالبة عن مدى التزامهم بحضور جميع المحاضرات أثناء تواجدهم داخل الحرم الجامعي، فوجد أن ١٢٤ طالبا و ١٠٥ طالبات من طلبة العينة يلتزمون بحضور جميع المحاضرات.

1- أوجد تقدير فترة ثقة 95% للفرق بين نسبتي الطلاب والطالبات الملتزمون بحضور المحاضرات.

2- عند مستوى معنوية 5%، هل تعطي هذه المعلومات دلالة كافية على أن الطالبات أكثر التزاما من الطلاب بحضور المحاضرات؟

الحل:

	طالبات	طلاب
النسبة في المجتمع	θ_1	θ_2
حجم العينة	ن$_1$ = ١٥٠	ن$_2$ = ٢٠٠
عدد الطلبة المواظبين	ك$_1$ = ١٠٥	ك$_2$ = ١٢٤

النسبة في العينة ق$_1$ = $\frac{١٠٥}{١٥٠}$ = ٠,٧٠ ق$_2$ = $\frac{١٢٤}{٢٠٠}$ = ٠,٦٢

النسبة المشتركة ق$_م$ = $\frac{١٠٥+١٢٤}{١٥٠+٢٠٠}$ = ٠,٦٥٤

تقدير فترة ثقة 95% للفرق بين النسبتين

$$Z_{1-\alpha\backslash 2} = Z_{0.975} = 1.96$$

(ق$_1$ − ق$_2$) ± $Z_{1-\alpha\backslash 2}$ $\sqrt{\frac{ق_1(1-ق_1)}{ن_1} + \frac{ق_2(1-ق_2)}{ن_2}}$

(٠,٧٠ − ٠,٦٢) ± ١,٩٦ × $\sqrt{\frac{٠,٧\times٠,٣}{١٥٠} + \frac{٠,٦٢\times٠,٣٨}{٢٠٠}}$

٠,٠٨ ± ٠,٠٩٩٥

الحد الأدنى = ٠,٠٨ − ٠,٠٩٩٥ = −٠,٠١٩٥

الحد الأعلى = ٠,٠٨ + ٠,٠٩٩٥ = ٠,١٧٩٥

2- هل الطالبات أكثر مواظبة على حضور المحاضرات؟

للإجابة على هذا السؤال نجري اختبارا للفروض يتضمن فيه فرض العدم أن الطالبات لسن أكثر التزاما وأكثر مواظبة على حضور المحاضرات (نسبة الإناث المواظبات على الحضور لا تختلف جوهريا عن نسبة الذكور، وقد تكون أقل). من جهة أخرى نضع الفرض البديل على أساس أن نسبة الإناث تزيد على نسبة الذكور. بالتالي وعلى ضوء تعريفنا للمجتمع الأول والمجتمع الثاني يكون الاختبار ذو طرف أيمن وتكون الفروض الإحصائية على الصورة:

$H_0: \theta_1 \leq \theta_2$ مقابل $H_1: \theta_1 > \theta_2$

أو

$H_0: \theta_1 - \theta_2 \leq$ صفر مقابل $H_1: \theta_1 - \theta_2 >$ صفر

إحصائية الاختبار:

$$= \frac{٠,٧٠ - ٠,٦٢}{\sqrt{٠,٦٥٤ (٠,٣٤٦) (\frac{١}{١٥٠} + \frac{١}{٢٠٠})}}$$

$$= \frac{٠,٠٨}{٠,٠٥١٤} = ١,٥٥.$$

القيمة الحرجة

اختبار طرف واحد عند مستوى معنوية $\alpha = 0.05$ من جدول النسب المختارة للتوزيع الطبيعي المعياري $Z_{1-\alpha} = Z_{0.95} = 1.645$

نتيجة الاختبار

بمقارنة قيمة إحصائية الاختبار بالقيمة الحرجة، نجد أن قيمة إحصائية الاختبار تقل عن القيمة الحرجة، ١,٥٥ > ١,٦٤٥. بالتالي لايمكننا رفض فرض العدم والذي يتضمن عدم وجود اختلاف جوهري بين النسبتين في المجموعتين . وذلك عند مستوى معنوية ٠,٠٥

٤-٥ إختبارات الفروض بتساوي تبايني مجتمعين مستقلين لهما توزيع طبيعي

رأينا في المبحث الثاني من هذا الفصل أن إحصائية اختبار الفرض بتساوي متوسطي مجتمعين مستقلين تعتمد على ما إذا كان تبايني المجتمعين متساويين أم لا، حيث نقوم في حالة تساوي التباينين بحساب تباين مشترك واحد كمتوسط مرجح لتبايني العينتين. ومن هذا تظهر أهمية دراسة اختبار للفروض يتعلق بتساوي تبايني مجتمعين طبيعيين ومستقلين، وهو ما نقوم به في هذا المبحث. ويشتمل أسلوب الاختبار على المراحل الأربع المعتادة لاختبارات الفروض.

الفروض الإحصائية

إذا رمزنا لتباين المجتمع الأول بالرمز σ_{1}^{2} ولتباين العينة المسحوبة منه بالرمز $ع_{1}^{2}$، وإذا رمزنا لتباين المجتمع الثاني بالرمز σ_{2}^{2}

ولتباين العينة المسحوبة منه بالرمز $ع_٢^٢$، تكون الفروض الإحصائية على أحدى الصورتين:

$$H_o: \sigma_١^٢ = \sigma_٢^٢ \quad \text{مقابل} \quad H_١: \sigma_١^٢ > \sigma_٢^٢$$

أو

$$H_o: \sigma_١^٢ = \sigma_٢^٢ \quad \text{مقابل} \quad H_١: \sigma_٢^٢ > \sigma_١^٢$$

ويمكن إعادة كتابة الفروض بصورة بديلة على أساس النسبة بين التباينين، وليس الفرق بينهما كما في حالتي اختبارات الفروض للفرق بين وسطين حسابيين أو الفرق بين نسبتين.

$$H_o: \frac{\sigma_١^٢}{\sigma_٢^٢} = ١ \quad \text{مقابل} \quad H_١: \frac{\sigma_١^٢}{\sigma_٢^٢} > ١$$

أو

$$H_o: \frac{\sigma_٢^٢}{\sigma_١^٢} = ١ \quad \text{مقابل} \quad H_١: \frac{\sigma_٢^٢}{\sigma_١^٢} > ١$$

يلاحظ من هذا أننا سوف نجري اختبار تساوي تبايني مجتمعين على صورة اختبار طرف أيمن. ويتم اختيار أحد الصورتين البديلتين للفروض بالنظر إلى تبايني العينتين بحيث يظهر التباين الأكبر دائما في البسط. فإذا وجدنا

أن $ع_1^2$ أكبر من $ع_2^2$، كتبنا الفرض البديل على الصورة $\frac{\sigma_1^2}{\sigma_2^2} > 1$. وإذا وجدنا أن $ع_2^2$ أكبر من $ع_1^2$، كتبنا الفرض البديل على الصورة $\frac{\sigma_2^2}{\sigma_1^2} > 1$.

إحصائية الاختبار

نقوم هنا باستخدام تباين كل عينة كمقدر نقطة لتباين المجتمع المسحوبة منه وتكون إحصائية الاختبار هي النسبة بين تبايني العينتين بحيث تكون هذه النسبة أكبر من الواحد الصحيح. وسوف نرمز لإحصائية الاختبار بالرمز F^* .

$$F^* = \frac{ع_1^2}{ع_2^2} \qquad \text{إذا كانت } ع_1^2 \text{ أكبر من } ع_2^2$$

$$F^* = \frac{ع_2^2}{ع_1^2} \qquad \text{إذا كانت } ع_2^2 \text{ أكبر من } ع_1^2.$$

وعندما يكون تباين كل عينة مقدر جيد لتباين المجتمع، أي يعبر عنه بصورة صادقة، فإن قيمة إحصائية الاختبار سوف تكون قريبة من الواحد الصحيح عندما يكون فرض العدم صحيحا. ومن ناحية أخرى في حالة عدم صحة فرض العدم، نتوقع أن يكون تباين العينة الذي يظهر في البسط أكبر بصورة جوهرية من تباين العينة الذي يظهر في المقام. ويتضمن هذا أن

تكون قيمة إحصائية الاختبار أكبر كثيرا من الواحد الصحيح.

القيمة الحرجة

لتحديد القيمة الحرجة التي تفصل بين منطقتي قبول ورفض فرض العدم، يلزم تحديد توزيع المعاينة لإحصائية الاختبار والتي تمثلها النسبة بين تبايني العينتين. ودون الخوض في مزيد من التفصيلات، التي تفوق مستوى الدارسين لهذا الكتاب، نقول بأن توزيع النسبة بين تباينين مستقلين محسوبين من بيانات عينات مسحوبة من مجتمعات طبيعية، وفي ظل صحة فرض العدم بتساوي التباينين، يكون توزيعا احتماليا يطلق عليه اسم توزيع النسبة F. ويكون لتوزيع F عددان لدرجات الحرية، درجات حرية البسط ($د_1$) ودرجات حرية المقام ($د_2$). وتوجد جداول لتوزيعات F كل منها يخص أحد قيم مستوى المعنوية شائعة الاستخدام. يوجد بنهاية هذا الكتاب جدول لمستوى المعنوية $\alpha = 0.05$ ، ويوجد جدول آخر لمستوى المعنوية $\alpha = 0.01$. ويشتمل الصف العلوي في الجدول على درجات حرية البسط ، بينما يشتمل العمود الأول على درجات حرية المقام. وعادة ما نكتب القيم المستخرجة من جداول توزيع F على الصورة $F_{(د_1 ، د_2 ، \alpha)}$. على سبيل المثال، يتم استخراج قيمة $F_{(8، 10، 0.05)}$ من الجدول الخاص بمستوى المعنوية 5% بالكشف أسفل درجات حرية البسط (8) وأمام درجات حرية المقام (10). ومن الجدول نجد أن:

$$F_{(8، 10، 0.05)} = ٣,٠٧$$

كذلك يتم استخراج قيمة $F_{(12، 6، 0.01)}$ من الجدول الخاص بمستوى المعنوية 1% بالكشف أسفل درجات حرية البسط (12) وأمام درجات حرية المقام (6). ومن الجدول نجد أن:

$$F_{(12، 6، 0.01)} = ٧,٧٢.$$

وفي اختبارات الفروض المتعلقة بتساوي تبايني مجتمعين، إذا كانت إحصائية الاختبار على الصورة $F^* = \frac{ع_1^2}{ع_2^2}$، تكون القيمة الحرجة هي

$F_{(ن_1-١، ن_2-١، \alpha)}$. وإذا كانت إحصائية الاختبار على الصورة

$F^* = \frac{ع_2^2}{ع_1^2}$، تكون القيمة الحرجة هي $F_{(ن_2-١، ن_1-١، \alpha)}$.

نتيجة الاختبار

إذا زادت قيمة إحصائية الاختبار عن القيمة الحرجة، نرفض فرض العدم ونقبل الفرض البديل ونستنتج أن المجتمع الذي ظهر تباين عينته في بسط إحصائية الاختبار يزيد بصورة جوهرية عن تباين المجتمع الآخر. وفي هذه الحالة عند إجراء اختبار تساوي متوسطي المجتمعين لا نقوم بحساب التباين المشترك. على الجانب الآخر، إذا قلت قيمة إحصائية الاختبار عن القيمة الحرجة، فإن هذا يعد دلالة على تساوي تبايني المجتمعين، وبالتالي

نقوم باستخدام التباين المشترك عند إجراء اختبار للفروض بتساوي متوسطي المجتمعين.

مثال ٤-٨

افترضنا في مثال ٤-٤ أن تبايني مجتمعي الذكور والإناث متساويين. والمطلوب اختبار صحة هذا الافتراض عند مستوى معنوية 1%.

الحل:

من بيانات مثال ٤-٤ نجد أن

	ذكور	إناث
تباين المجتمع	σ_1^2	σ_2^2
تباين العينة	$ع_1^2 = ١$	$ع_2^2 = ٠,٦٤$
حجم العينة	$ن_1 = ١٠$	$ن_2 = ٨$

حيث أن تباين عينة الذكور أكبر من تباين عينة الإناث، تكون الفروض الإحصائية على الصورة:

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad \text{مقابل} \quad H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1$$

إحصائية الاختبار

$$F^* = \frac{ع_1^2}{ع_2^2}$$

$$= \frac{١}{٠,٦٤}$$

$$= ١,٥٦٢٥$$

يتم استخراج القيمة الحرجة من جدول توزيع F الخاص بمستوى المعنوية 1% بالكشف أسفل درجات حرية البسط وهي حجم العينة الأولى مطروحا منه واحد ($ن_1 - 1 = 9$) (لأنها ظهرت في بسط الإحصائية) وأمام درجات حرية المقام وهي حجم العينة الثانية مطروحا منه واحد ($ن_2 - 1 = 7$).

$F_{(9, 7, 0.01)}$ = ٦,٧٢.

نتيجة الاختبار

حيث أن قيمة إحصائية الاختبار تقل كثيرا عن القيمة الحرجة، فإننا نقبل فرض العدم ونستنتج أنه يمكن أن يكون لمجتمعي توزيع درجات الذكور وتوزيع درجات الإناث تباين متساوي، أو لا يوجد فرق جوهري بين التباينين. لذلك عند إجراء اختبار للفروض يتعلق بمتوسطي المجتمعين، قمنا باستخدام التباين المشترك.

تمـــارين الفصل الرابع

(1) سحبت عينة عشوائية حجمها ١٢ مفردة من مجتمع له توزيع طبيعي ع(μ_1، ٢٤) ، كما سحبت عينة عشوائية حجمها ٢٠ مفردة من مجتمع آخر، مستقل عن المجتمع الأول وله توزيع طبيعي ع(μ_2، ١٥) .

1- اكتب مقدر نقطة للفرق بين متوسطي المجتمعين.

2- ما هو توزيع المعاينة لمقدر النقطة في المطلوب السابق؟

3- ما هو احتمال أن يكون خطأ تقدير مقدر النقطة للفرق بين متوسطي المجتمعين هو ثلاث وحدات على الأكثر.

4- على فرض أن متوسط المجتمع الأول يزيد عن متوسط المجتمع الثاني بثلاث وحدات، ما هو ح(| $\overline{س}_1$ - $\overline{س}_2$ | ≥ ٢)؟

5- نظرا لكبر حجم المجتمع الثاني وكبر أهميته النسبية نريد اختيار عينتين أخرتين من المجتمعين بحيث يكون حجم عينة المجتمع الأول نصف حجم عينة المجتمع الثاني، ما هو حجم كل من العينتين الذي يجعل أقصى خطأ لتقدير الفرق بين

متوسطي المجتمعين باستخدام الفرق بين متوسطي العينتين وحدة واحدة باحتمال 95%.

(2) لقياس تأثير الموقع على حجم مبيعات المحلات التجارية، تم اختيار عينة عشوائية من ١٠ محلات للملابس الجاهزة تقع جميعها داخل مراكز للتسوق وعينة أخرى من ١٥ محلا منتشرة في الأحياء السكنية وتم تسجيل مبيعات كل محل منها على مدار أسبوع، فوجد أن متوسط حجم الميعات اليومية خلال فترة الدراسة لمحلات مراكز التسوق هو ٣٤٥٠ جنيها بانحراف معياري ٨٠ جنيها. كذلك وجد أن متوسط المبيعات اليومية للمجموعة الأخرى هو ٣١٢٥ جنيها بانحراف معياري ٦٠ جنيها. على فرض أن قيم المبيعات اليومية لمحلات الملابس الجاهزة تتبع توزيع طبيعي تقريبا وبنفس التباين فالمطلوب:

1- عند مستوى معنوية 5%، اختبر تأثير الموقع على حجم المبيعات.

2- اكتب تقدير فترة ثقة 95% للفرق بين متوسطي قيم المبيعات اليومية لمحلات الملابس الجاهزة في كلا الموقعين.

(3) تم تدريب عينة من ٢٠ رجلا و٢٥ سيدة على إنجاز مهمة معينة، ثم طلب من كل منهم القيام بإنجاز تلك المهمة وتم قياس الزمن فكان متوسط الزمن للرجال هو ٥ دقائق وللسيدات ٤ دقائق. وإذا علمت أن

توزيع زمن إنجاز هذه المهمة يتبع بصفة عامة توزيع طبيعي تقريبا بانحراف معياري متساوي قدره ٧٢ ثانية لكلا الرجال والسيدات.

1- عند مستوى معنوية 1% هل تؤيد هذه البيانات أن السيدات أكثر كفاءة من الرجال في إنجاز المهمة؟

2- أوجد تقدير فترة ثقة 98% للفرق بين متوسطي زمن الرجال والسيدات.

(4) لاختبار تأثير نوع جديد من أدوية علاج مرض ارتفاع ضغط الدم، تم اختيار عينة من ٢٠ مريضاً في نفس الحالة الصحية ونفس السن تقريباً وتم تقسيمهم إلى مجموعتين إحداهما تتكون من ١٢ مريضاً تم إعطائهم أقراصاً خالية من المادة الفعالة بينما تم إعطاء الثمانية المتبقين الدواء الجديد. وبعد فترة من تناول العلاج تم قياس ضغط الدم لكل من المرضى فكانت بيانات بسط الضغط كما يلي:

المجموعة الأولى: ١٨٠، ١٩٠، ١٧٥، ١٩٠، ٢١٠، ١٨٤، ١٧٣، ١٨٠، ١٨٣، ٢٠٠، ١٩٥، ١٦٥.

المجموعة الثانية: ١٤٠، ١٦٥، ١٥٠، ١٥٥، ١٤٠، ١٨٠، ١٧٠، ١٥٠.

وعلى فرض أن توزيع ضغط الدم لدى المرضى هو توزيع طبيعي بنفس التباين للمجموعتين

أ) علق على فاعلية الدواء الجديد عند مستوى معنوية 1%.

ب) اكتب تقدير فترة ثقة 99% للفرق بين متوسطي مقياس الضغط للمرضى الذين لا يتناولون أي علاج والمرضى الذين يتناولون الدواء الجديد.

(5) لقياس مدى فاعلية نظام جديد لإنقاص الوزن في أحد المعاهد الرياضية، تم تحديد أوزان ثمانية أشخاص قبل تطبيق هذا النظام وتحديد أوزانهم بعد مضي شهرين على تطبيق هذا النظام فكانت البيانات كما يلي:

الوزن قبل النظام: ٨٠، ٧٨،٨٥ ، ٨٤، ٨٧، ٩٦، ٩٢، ٨٤

الوزن بعد النظام: ٧٧، ٨٤، ٧٦، ٨١، ٨٢، ٩٢، ٩٠، ٨٠

أ) اختبر معنوية تأثير النظام على إنقاص الوزن عند مستوى معنوية 5%.

ب) أوجد تقدير فترة ثقة 95% للفرق بين متوسطي الوزن قبل وبعد تطبيق النظام.

(6) انتجت إحدى شركات البترول مادة جديدة وادعت أنها إذا ما أضيفت إلى وقود السيارة فإنها تزيد من عدد الكيلومترات التي تقطعها السيارة لكل لتر مستهلك من الوقود. ولتجربة تأثير هذه المادة تم اختيار عينة عشوائية من ست سيارات وتم تسييرها لمدة أسبوع باستخدام وقود عادي وبدون إضافة المادة، ثم تم تسيير نفس السيارات لمدة أسبوع

آخر مع إضافة المادة لنفس نوع الوقود. وقد سجلت البيانات التالية لمتوسط عدد الكيلومترات التي قطعتها كل سيارة لكل لتر خلال الأسبوعين:

بدون المادة: ٥,٧٥ ، ٦,٢٥ ، ٥ ، ٦ ، ٥,٥ ، ٧

بإضافة المادة: ٦,٢٥ ، ٦ ، ٥,٥ ، ٧ ، ٦,٥ ، ٨,٢٥

أ) هل تنصح قائدي السيارات بإضافة المادة الجديدة عند مستوى معنوية 5%؟

ب) أوجد تقدير فترة ثقة 95% للفرق بين متوسطي عدد الكيلومترات لكل لتر عند إضافة المادة وعند عدم إضافتها.

(7) لدى إحدى الشركات سلسلة مطاعم في مدينتي (أ ، ب) وفي عينة من ٢٠٠ عميل في المدينة (أ) وجد أن هناك ١٢٠ شخصاً يعتقدون أن مستوى الخدمة ممتاز. بينما في عينة من ٢٥٠ عميل في المدينة (ب) وجد أن عدد العملاء الذين يعطون نفس التقدير للخدمة هو ١٢٥.

أولاً: هل تعطي هذه المعلومات دلالة على عدم وجود اختلاف بين مستوى خدمة المطاعم في المدينتين من وجهة نظر العملاء عند مستوى معنوية 5%؟

ثانياً: اكتب تقدير فترة ثقة 95% للفرق بين نسبتي العملاء الذين يعتقدون أن مستوى الخدمة ممتاز في المدينتين.

(8) في عينة من ٤٠٠ طالب و ٢٠٠ طالبة في الجامعة وجد أن عدد من استخدم شبكة المعلومات **The Internet** مرة واحدة على الأقل هو ١١٢ طالباً و٤٦ طالبة.

اختبر الفرض بأن نسبة مستخدمي الشبكة من الطلاب أعلى من نسبة الطالبات عند مستوى معنوية 5% ، ثم أوجد تقدير فترة ثقة 95% للفرق بين النسبتين.

الفصل الخامس

تحليل التباين

Analysis of Variance

(٥ ـــ ١) مقدمة :

سنتعرض في هذا الفصل لكيفية المقارنة بين أكـــثر مــن متوســط مجتمعين باستخدام البيانات التي تُجمع وفقاً لتصميم تجريبي معين .

وقد كان للتجارب الزراعية دوراً رائداً في مجال تصميــــم التجــارب Experimental design لذلك نجد أن مصطلحات هذا الموضوع يغلب عليـــه الطابع الزراعي . فمعظم التجارب الزراعية تتضمن معالجة الوحدات التجريبية بطريقتين أو أكثر ثم المقارنة بين متوسطات المجتمعات المناظرة للمعالجــــات المختلفة . على سبيل المثال المقارنة بين متوسطات إنتاجية الفدان في مجموعة من حقول القمح تم زراعتها باستخدام أربعة أنواع مختلفة مـــن الأســمدة . أو المقارنة بين متوسطات الزيادة في أوزان مجموعة من الأبقار الناتجة عن اتباع أربعة أنواع من أنظمة التغذية للتسمين . ويطلق على كل من أنواع الأســـمدة . أنواع أنظمة التغذية للتسمين . اسم المعالجات . والهدف من مثل هذه التجــارب المقارنة بين متوسطات المجتمعات المناظرة للأربع أنواع من المعالجات .

ومن مميزات جمع البيانات وفقاً لتصميم تجريبـــي أنـــه يمكننـــا مــن الحصول على قدر من المعلومات أكبر من ذلك الذي يتم الحصــــول عليـــه إذا

لم تكن البيانات قد جمعت وفقاً لتصميم تجريبي . كما أنه يمكن الباحث من تحليل البيانات بأسلوب بسيط يعرف بتحليل التباين .

وسوف نتناول في هذا الفصل تحليل التباين لتصميم تام العشوائية ، ولتصميم القطاعات الكاملة العشوائية .

(٥ ـ ٢) المقارنة بين أكثر من متوسطي مجتمعين :

التصميم التام العشوائية :

Completely randomized design:

التصميم التام العشوائية هو ذلك التصميم الذي يتم فيه اختيار عينات عشوائية مستقلة من كل مجتمع من المجتمعات تحت الدراسة . وباستخدام هذه العينات يمكن المقارنة بين متوسطات هذه المجتمعات ، ويوضح الجدول (٥ - ١) الرموز المتعلقة بالتصميم التام العشوائية وذلك على فرض أننا نريد المقارنة بين متوسطات ل من المجتمعات .

وتتم المقارنة بين متوسطات المجتمعات (المعالجات) μ_1 ، μ_2 ... ، $\mu_ل$ لتقرير ما إذا كان هناك فرق بينها من خلال دراسة الاختلاف (التباين) بين الأوساط الحسابية للعينات . فالإختلاف الأكبر دليل أقوى على وجود فرق بين μ_1 ، μ_2 ... ، $\mu_ل$ ويقاس هذا الاختلاف بالمجموع المرجح لمربعات انحرافات الأوساط الحسابية للعينات $\overline{س}_1$ ، $\overline{س}_2$ ، ... ، $\overline{س}_ل$ عن الوسط العام $\overline{\overline{س}}$ ويعرف هذا المجموع بمجموع المربعات بين متوسطات العينات أو مجموع المربعات بين المعالجات . ويمكن التعبير عنه كالآتي :

$$\text{مجموع المربعات بين المعالجات} = \sum_{ر=1}^{ل} ن_ر \left(\overline{س}_ر - \overline{\overline{س}} \right)^2$$

جدول (٥-١)

الرموز المستخدمة في التصميم التام العشوائية

<table>
<tr><td></td><td>المجتمعات (المعالجات)</td></tr>
<tr><td>المتوسط
التباين</td><td>١ ٢ ٣ ... ل
μ_1 μ_2 μ_3 $\mu_ل$
σ^2_1 σ^2_2 σ^2_3 $\sigma^2_ل$</td></tr>
<tr><td></td><td>العينات العشوائية المستقلة</td></tr>
<tr><td>حجم العينة
مجموع المشاهدات
الوسط الحسابي للعينة</td><td>١ ٢ ٣ ... ل
$ن_1$ $ن_2$ $ن_3$ ... $ن_ل$
$م_1$ $م_2$ $م_3$ ... $م_ل$
$\bar{س}_1$ $\bar{س}_2$ $\bar{س}_3$ ... $\bar{س}_ل$</td></tr>
<tr><td colspan="2">العدد الكلي للمشاهدات = ن = $ن_1 + ن_2 + ... ن_ل$
مجموع ن من المشاهدات = $\sum_{و=1}^{ن} س_و$
الوسط العام = $\bar{\bar{س}}$
مجموع مربعات ن من المشاهدات = $\sum_{و=1}^{ن} س_و^2$</td></tr>
</table>

ويلاحظ أنه تم ترجيح مربع كل فرق بين الوسط الحسابي للعينة والوسط العام بحجم العينة المناظرة كما يلاحظ أيضاً أن هذا المجموع سيكون كبيراً إذا كان الاختلاف كبيراً بين الأوساط الحسابية للعينات . فبفرض أننا نريد اختبار فرض العدم بتساوي متوسطات ل من المعالجات ، أي أن :

$$H_0 : \mu_1 = \mu_2 = ... = \mu_ل$$

في مقابل الفرض البديل :

H_1 : يوجد فرق بين متوسطين على الأقل .

في هذه الحالة نجد أن القيم الكبيرة لمجموع المربعات بين المعالجات هي التي ستؤيد الفرض البديل . بمعنى أنه إذا كان مجموع مربعات الفروق بين الأوساط الحسابية للعينات والوسط العام كبيراً فسيكون هناك اتجاهاً لتأييد الفرض بوجود فرق بين متوسطات المجتمعات .

والآن نتساءل إلى أي مدى يمكن اعتبار مجموع المربعات بين المعالجات كبيراً قبل رفض فرض العدم وقبول البديل ؟ تتوقف الإجابة على مقارنة هذا المجموع كمقياس للاختلاف بين متوسطات العينات بمقياس آخر للاختلاف داخل العينات .

ويمكن قياس الاختلاف داخل العينات بإيجاد مجموع مشترك لمجاميع مربعات انحرافات المشاهدات داخل العينة عن وسطها الحسابي ويعرف هذا المجموع بمجموع المربعات داخل العينات أو ، بمجموع مربعات الخطأ حيث أنه يقيس الاختلاف الغير مفسر بواسطة الفروق بين متوسطات العينات ويمكن التعبير عن هذا المجموع بالصورة الآتية :

مجموع مربعات الخطأ = $\sum_{و=1}^{ن_1} (س_{1و} - \bar{س}_1)^2 + \sum_{و=1}^{ن_2} (س_{2و} - \bar{س}_2)^2$

$$+ \;\ldots\ldots\ldots + \sum_{و=1}^{ن_ل} (س_{لو} - \bar{س}_ل)^2$$

$$= \sum_{ر=1}^{ل} \sum_{و=1}^{ن_ر} (س_{رو} - \bar{س}_ر)^2$$

حيث تشير $س_{رو}$ إلى المشاهدة رقم و في العينة رقم ر

ولمقارنة الاختلاف بين متوسطات العينات بالاختلاف داخل العينات نتبع ما يلي :

١ ـ إيجاد متوسط المربعات بين المعالجات بقسمة مجموع المربعات بين المعالجات على درجات الحرية المرتبطة به (ل – ١) وهي عبارة عن درجة واحدة لمتوسط كل معالجة من ل من المعالجات مطروحاً منها الواحد الصحيح حيث تم تقدير الوسط العام أي أن :

$$\text{متوسط المربعات بين المعالجات} = \frac{\text{مجموع المربعات بين المعالجات}}{\text{ل} - ١}$$

٢ ـ إيجاد متوسط مربعات الخطأ بقسمة مجموع مربعات الخطأ على درجات الحرية المرتبطة به (ن – ل) وهي عبارة عن درجة لكل مشاهدة من ن من المشاهدات مطروحاً منها ل (عدد متوسطات العينات التي يتم تقديرها) :

$$\text{متوسط مربعات الخطأ} = \frac{\text{مجموع مربعات الخطأ}}{\text{ن} - \text{ل}}$$

٣ ـ حساب إحصائية ف :

$$\text{ف} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

حيث تشير القيم الكبيرة لإحصائية ف إلى وجود فروق كبيرة بين متوسطات العينات وبالتالي تأييد الفرض البديل بوجود فرق بين متوسطات المجتمعات . ويلاحظ أن إحصائية ف تتضمن المقارنة بين مصدرين للاختلاف (التباين) أحدهما يرجع إلى الفروق بين متوسطات العينات والآخر يرجع إلى الفروق داخل العينات ولذلك سُمي هذا الأسلوب لمقارنة متوسطات المجتمعات بتحليل التباين Analysis of Variance (ANOVA) .

ويمكن تلخيص عناصر اختبار الفرق بين متوسطات ل من المجتمعات والشروط الواجب توافرها كالآتي :

اختبار الفرق بين متوسطات ل من المعالجات باستخدام التصميم التام العشوائية
الفروض الإحصائية : H_0 : $\mu_1 = \mu_2 = \cdots\cdots = \mu_{ل}$ H_1 : يوجد فرق بين متوسطين على الأقل إحصائية الاختبار : ف = $\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$ الشروط : ١ ــ يتبع كل مجتمع من المجتمعات المراد المقارنة بين متوسطاتها لتوزيع معتدل ٢ ــ تباينات المجتمعات متساوية . ٣ ــ العينات عشوائية ومستقلة . منطقة الرفض : ف > ف$_{(\alpha ، ح_1 ، ح_2)}$ حيث $ح_1$ = ل − ١ = عدد درجات الحرية المرتبطة بمتوسط المربعات بين المعالجات ، $ح_2$ = ن − ل = عدد درجات الحرية المرتبطة بمتوسط مربعات الخطأ .

وبالرغم من إمكانية استخدام الصيغ السابقة لحساب مجموع المربعات بين المعالجات ومجموع مربعات الخطأ إلا أنه يوجد صيغ أخرى أكثر بساطة في الحساب . بالإضافة إلى ذلك إمكانية اللجوء إلى فكرة تجزئة مجموع مربعات انحرافات كل المشاهدات عن الوسط العام ويعرف هذا المجموع بمجموع المربعات الكلي . أي أن :

$$\text{مجموع المربعات الكلي} = \sum_{و=1}^{ن} (س_و - \bar{\bar{س}})^2$$

= مجموع المربعات بين المعالجات + مجموع مربعات الخطأ

والفائدة التي تعود علينا من وراء فكرة هذه التجزئة هي إمكانية إيجــاد مجموع مربعات الخطأ كمتمم حسابي أي أن :

مجموع مربعات الخطأ = مجموع المربعات الكلي – مجموع المربعات بين المعالجات

صيغ الحسابات اللازمة لتحليل التباين للتصميم التام العشوائية

معامل التصحيح = $\frac{(\text{مجموع كل المشاهدات})^2}{\text{العدد الكلي للمشاهدات}}$

$= \frac{(\text{مجـ س})^2}{\text{ن}}$

مجموع المربعات الكلي = مجموع مربعات المشاهدات – معامل التصحيح

= مجـ س^2 – معامل التصحيح

مجموع المربعات بين المعالجات = $\left(\begin{array}{c}\text{مجموع مربعات مجاميع المعالجات}\\ \text{مقسوماً كل منها على عدد}\\ \text{المشاهدات للمعالجة}\end{array}\right)$ – معامل التصحيح

$= \frac{\text{م}_1^2}{\text{ن}_1} + \frac{\text{م}_2^2}{\text{ن}_1} + ... + \frac{\text{م}_\text{ل}^2}{\text{ن}_\text{ل}}$ – معامل التصحيح

مجموع مربعات الخطأ = مجموع المربعات الكلي – مجموع المربعات بين المعالجات

متوسط المربعات بين المعالجات = $\frac{\text{مجموع المربعات بين المعالجات}}{\text{ل} - 1}$

متـــــوسط مربعــــــات الخطـــأ = $\frac{\text{مجموع مربعات الخطأ}}{\text{ن} - \text{ل}}$

إحصـــــائيـــة الاختبــــار : ف = $\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$

وعادة ما يتم تلخيص نتائج الحسابات في صورة جدول يعرف بجدول تحليل التباين . والجدول (٥ ــ ٢) يوضح جدول تحليل التباين لتصميم تام العشوائية . ويوضح فيه مصدر الاختلاف ، درجات الحرية ، مجموع المربعات ، متوسط المربعات وإحصائية الاختبار ف .

جدول (٥ ــ ٢)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	ل ــ ١	مجموع المربعات بين المعالجات	متوسط مربعات المعالجات	متوسط مربعات المعالجات / متوسط مربعات الخطأ
الخطأ	ن ــ ل	مجموع مربعات الخطأ	متوسط مربعات الخطأ	
الكلي	ن ــ ١	مجموع مربعات الكلي		

مثال (١) :

أرادت إحدى المؤسسات الائتمانية مقارنة متوسطات الديون المستحقة الدفع على العملاء في ثلاث مستويات مختلفة للدخل السنوي بالجنيهات (أقل من ١٢٠٠٠ ، ١٢٠٠٠ ــ ٢٥٠٠٠ ، أكبر من ٢٥٠٠٠) فقامت باختيار عينة مكونة من حسابات عشرة عملاء من كل مجموعة دخلية ، وتم تسجيل الديون المستحقة الدفع على كل عميل وجدول (٥ ــ ٣) يلخص النتائج التي تم الحصول عليها . المطلوب استخدام هذه البيانات لاختبار الفرض بأن متوسطات الديون المستحقة الدفع في المجموعات الدخلية الثلاثة متساوية في مقابل الفرض البديل بأنها مختلفة . استخدم مستوى معنوية ٠,٠٥ .

جدول رقم (٥ ــ ٣)

الديون المستحقة الدفع في المجموعات الدخلية المختلفة

أقل من ١٢٠٠٠	١٢٠٠٠ – ٢٥٠٠٠	٢٥٠٠٠ فأكثر
١٤٨	٥١٣	٣٣٥
٧٦	٢٦٤	٦٤٣
٣٩٣	٤٣٣	٢١٦
٥٢٠	٩٤	٥٣٦
٢٣٦	٥٣٥	١٢٨
١٣٤	٣٢٧	٧٢٣
٥٥	٢١٤	٢٥٨
١٦٦	١٣٥	٣٨٠
٤١٥	٢٨٠	٥٩٤
١٥٣	٣٠٤	٤٦٥
٢٢٩٦	٣٠٩٩	٤٢٧٨

الحـــــل :

حيث أننا نريد أن نختبر فرض العــدم بتســاوي متوســطات الديــون المستحقة الدفع في المجموعات الدخلية الثلاثة ، وبفـــرض أن μ_1 ، μ_2 ، μ_3 تمثل على الترتيب متوسط الديون المستحقة الدفـــع فـــي المجموعـــة الدخليـــة المنخفضة والمتوسطة والمرتفعة المستوى فتكون عناصر الاختبار كالآتي :

الفروض الإحصائية :

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : يوجد فرق بين متوسطين على الأقل

إحصائية الاختبار : ف = $\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$

الشروط : ١ ــ تتوزع الديون المستحقة الدفع في كل مجموعة دخلية توزيعاً معتدلاً .

٢ ــ تساوي التباين لتوزيعات الديون المستحقة الدفع في المجموعات الدخلية الثلاثة .

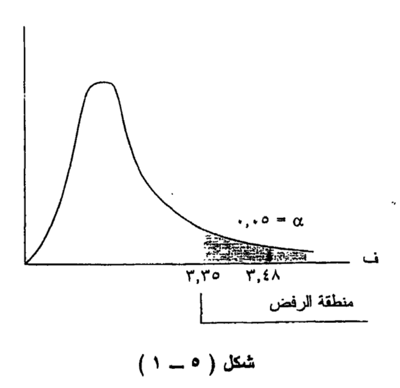
٣ ــ العينات عشوائية ومستقلة عن بعضها البعض .

منطقة الرفض : ف > ف$_{(\alpha ، ٧_١ ، ٧_٢)}$

حيث $٧_١$ = ل – ١ = ٣ – ١ = ٢

، $٧_٢$ = ن – ل = ٣٠ – ٣ = ٢٧

ف$_{(\alpha ، ٧_١ ، ٧_٢)}$ = ف$_{(٠,٠٥ ، ٢ ، ٢٧)}$ = ٣,٣٥ (أنظر الشكل ٥ ــ ١)



شكل (٥ ــ ١)

من جدول (٥ ــ ٣) نجد أن :

$م_١ = ٢٢٩٦$ ، $م_٢ = ٣٠٩٩$ ، $م_٣ = ٤٢٧٨$

$مجـ س = م_١ + م_٢ + م_٣ = ٩٦٧٣$

$مجـ س^٢ = (١٤٨)^٢ + (٧٦)^٢ + ... + (٤٦٥)^٢ = ٤٠٨٨٣٤١$

$معامل\ التصحيح = \frac{(مجـ س)^٢}{ن} = \frac{(٩٦٧٣)^٢}{٣٠} = ٣١١٨٨٩٧,٦$

مجموع المربعات الكلي = مجـ س^٢ – معامل التصحيح

$= ٤٠٨٨٣٤١ - ٣١١٨٨٩٧,٦$

$= ٩٦٩٤٤٣,٤$

$مجموع\ المربعات\ بين\ المعالجات = \frac{م_١^٢}{ن_١} + \frac{م_٢^٢}{ن_٢} + \frac{م_٣^٢}{ن_٣} - معامل\ التصحيح$

$= \frac{(٢٢٩٦)^٢}{١٠} + \frac{(٣٠٩٩)^٢}{١٠} + \frac{(٤٢٧٨)^٢}{١٠} - ٣١١٨٨٩٧,٦$

$= ٣٣١٧٦٧٠,١ - ٣١١٨٨٩٧,٦ = ١٩٨٧٧٢,٥$

مجموع مربعات الخطأ = مجموع المربعات الكلي – مجموع المربعات بين المعالجات

$= ٩٦٩٤٤٣,٤ - ١٩٨٧٧٢,٥ = ٧٧٠٦٧٠,٩$

$متوسط\ المربعات\ بين\ المعالجات = \frac{مجموع\ المربعات\ بين\ المعالجات}{ل - ١}$

$= \frac{١٩٨٧٧٢,٥}{٢} = ٩٩٣٨٦,٢٥$

$متوسط\ مربعات\ الخطأ = \frac{مجموع\ مربعات\ الخطأ}{ن - ل}$

$= \frac{٧٧٠٦٧٠,٩}{٢٧} = ٢٨٥٤٣,٤$

$$\text{ف} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

$$= \frac{٩٣٣٨٦,٢٥}{٢٨٥٤٣,٤} = ٣,٤٨$$

وحيث أن قيمة ف المحسوبة تقع في منطقة الرفض كما هو مبين بالشكل (٥ ــ ١) فإننا نستنتج أن هناك اختلاف بين متوسط الديون المستحقة الدفع في مجموعتين على الأقل من المجموعات الدخلية وذلك عند مستوى معنوية ٠,٠٥ وجدول (٥ ــ ٤) يوضح جدول تحليل التباين لمثالنا الحالي :

جدول (٥ ــ ٤)

جدول تحليل التباين لمثال (١)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٢	١٩٨٧٧٢,٥	٩٩٣٨٦,٢٥	٣,٤٨
الخطأ	٢٧	٧٧٠٦٧٠,٩	٢٨٥٤٣,٤	
الكلي	٢٩	٩٦٩٤٤٣,٤		

ونظرا لما يتضمنه التصميم التام العشوائية من اختيار لعينات عشوائية مستقلة فإنه يمكن إيجاد فترة ثقة لمتوسط المعالجة الواحدة وكذلك للفرق بين متوسطي معالجتين مع الأخذ في الاعتبار أن يستخدم متوسط مربعات الخطأ كمقدر للتباين σ^2 أي أن

$$ع^2 = \text{متوسط مربعات الخطأ} = \frac{\text{مجموع مربعات الخطأ}}{ن - ل}$$

وبناءا على ذلك يمكن تلخيص كيفية إيجاد فترة ثقة لمتوسط المعالجة الواحدة وكذلك للفرق بين متوسطي معالجتين كالآتي :

فترة ثقة ١٠٠ (١ − α) ٪ لمتوسط المعالجة ر

$$\bar{س}_ر \pm ت_{(\frac{\alpha}{2}، ن - ل)} \frac{ع}{\sqrt{ن_ر}}$$

فترة ثقة ١٠٠ (١ − α) ٪ للفرق بين متوسطي المعالجتين ر ، ز

$$(\bar{س}_ر - \bar{س}_ز) \pm ت_{(\frac{\alpha}{2}، ن - ل)} ع \sqrt{\frac{1}{ن_ر} + \frac{1}{ن_ز}}$$

حيث : ع = $\sqrt{\text{متوسط مربعات الخطأ}}$

مثال (٢) :

في المثال (١) المطلوب إيجاد فترة ثقة ٩٥% لمتوسط الديون المستحقة الدفع لمجموعة العملاء الذين يقل دخلهم السنوي عن ١٢٠٠٠ جنيها .

الحل :

من جدول تحليل التباين (٥ ــ ٤) نجد أن متوسط مربعات الخطأ = ٢٨٥٤٣,٤ وبالتالي فإن :

ع = $\sqrt{\text{متوسط مربعات الخطأ}}$

= $\sqrt{٢٨٥٤٣,٤}$ = ١٦٨,٩

كما يمكن إيجاد الوسط الحسابي لعينة الديون المستحقة الدفع لمجموعة العملاء الذين يقل دخلهم عن ١٢٠٠٠ جنيها كالآتي :

$$\overline{\text{س}_{١}} = \frac{\text{مج}_{١}}{\text{ن}_{١}} = \frac{٢٢٩٦}{١٠} = ٢٢٩,٦$$

وبناءا على ذلك تكون فترة ثقة ٩٥٪ لمتوسط الديون المستحقة الدفع لمجموعة العملاء ، الذين يقل دخلهم عن ١٢٠٠٠ جنيها على الصورة :

$$\overline{\text{س}_{١}} \pm \text{ت}_{(\frac{\alpha}{٢} ، \text{ن} - \text{ل})} \frac{\text{ع}}{\sqrt{\text{ن}_{١}}}$$

$$= \overline{\text{س}_{١}} \pm \text{ت}_{(٢٧٠,٠٢٥)} \frac{\text{ع}}{\sqrt{\text{ن}_{١}}}$$

$$= ٢٢٩,٦ \pm ٢,٠٥٢ \left(\frac{١٦٨,٩}{\sqrt{١٠}} \right)$$

$$= ٢٢٩,٦ \pm ١٠٩,٦$$

$$= (١٢٠ ، ٣٣٩,٢)$$

ومن الملاحظ أن هذه الفترة تتسم بالاتساع ويرجع السبب في ذلك إلى زيادة درجة التفاوت بين مفردات العينة ، فهي تتفاوت من ٥٥ جنيها إلى ٢٥٠ جنيها ولكن إذا أردنا الحصول على تقدير دقيق لمتوسط المعالجة بفترة ثقة أضيق من تلك الفترة فيجب تزويد حجم العينة .

مثال (٣) :

في المثال (١) أوجد فترة ثقة ٩٥٪ للفرق بين متوسطي الديون المستحقة الدفع على العملاء الذين يقل دخلهم السنوي عن ١٢٠٠٠ جنيها والذين يزيد دخلهم السنوي عن ٢٥٠٠٠ جنيها .

الحل :

من المثال (١) يمكن إيجاد الوسط الحسابي لعينة الديون المستحقة الدفع على العملاء الذين يزيد دخلهم عن ٢٥٠٠٠ جنيها في السنة كالآتي :

$$\overline{س_٣} = \frac{م_٣}{ن_٣} = \frac{٤٢٧٨}{١٠} = ٤٢٧,٨$$

ومن المثال (٢) وجدنا أن $\overline{س_١}$ = ٢٢٩,٦ . وبالتالي تكون فترة ثق
٩٥٪ للفرق ($\mu_٣ - \mu_١$) على الصورة :

$$(\overline{س_٣} - \overline{س_١}) \pm ت_{(\frac{\alpha}{٢}، ن - ل)} ع \sqrt{\frac{١}{ن_٣} + \frac{١}{ن_١}}$$

$$= (٤٢٧,٨ - ٢٢٩,٦) \pm (٢,٠٥٢)(١٦٨,٩)\sqrt{\frac{١}{١٠} + \frac{١}{١٠}}$$

$$= ١٩٨,٢ \pm ١٥٥,٠$$

$$= (٤٣,٢ ، ٣٥٣,٢)$$

ومن الملاحظ أن هذه الفترة للفرق بين المتوسطين ($\mu_٣ - \mu_١$) تـ
بالاتساع الشديد كما هو الحال بالنسبة لفترة الثقة للمتوسط الواحد ويرجع السب
في ذلك أيضا إلى التفاوت الكبير بين المفردات داخل العينة الواحدة ولذلك حتـ
نستطيع الحصول على فترة أضيق لابد من تزويد حجم العينات الثلاثـ
ويلاحظ أيضا أن هذه الفترة تحتوي على قيم موجبة فقط مما يجعلنا نثق بدر
٩٥٪ في أن يكون متوسط الديون المستحقة الدفع على العملاء الذين يزيد دخـ
السنوي عن ٢٥٠٠٠ جنيها أعلى من نظيره بالنسبة للعملاء الذين يقل دخلـ
السنوي عن ١٢٠٠٠ جنيها .

(٥ ــ ٣) تصميم القطاعات الكاملة العشوائية :
Completely Randomized Block Design:

يستخدم تصميم القطاعات الكاملة العشوائية مجموعات من الوحدات التجريبية المتجانسة للمقارنة بين متوسطات المجتمعات المتعلقة بعدد من المعالجات . ففي هذا النوع من التصميمات يتم تقسيم الوحدات التجريبية إلى مجموعات بشرط أن تكون هذه الوحدات داخل المجموعة الواحدة متجانسة ومساوية في عددها لعدد المعالجات . كما يشترط أن يكون توزيع المعالجات داخل المجموعة عشوائيا . ويطلق على كل مجموعة من هذه المجموعات اسم قطاع BLOCK . ويوضح الجدول (٥ ــ ٥) الرموز التي ستستخدم عند تحليل نتائج تصميم القطاعات الكاملة العشوائية . وذلك بفرض أن لدينا ط من القطاعات وأن عدد المعالجات المطلوب المقارنة بين متوسطاتها يساوي ل .

جدول (٥ ــ ٥)

الرموز المستخدمة في تصميم القطاعات الكاملة العشوائية

	المعالجة (١)	المعالجة (٢)	......	المعالجة (ل)
حجم العينة	ط	ط	......	ط
مجموع المشاهدات	$م_{١}$	$م_{٢}$		$م_{ل}$
	قطاع (١)	قطاع (٢)		قطاع (ط)
حجم العينة	ل	ل	......	ل
مجموع المشاهدات	$ق_{١}$	$ق_{٢}$		$ق_{ط}$

العدد الكلي للمشاهدات = ل ط = ن

مجموع ن من المشاهدات = مجـ س

مجموع ن من مربعات المشاهدات = مجـ $س^{٢}$

وهدفنا الآن هو استخدام تصميم القطاعات الكاملة العشوائية لاختبار فرض العدم بتساوي متوسطات المعالجات في مقابل البديل بوجود فرق بين هذه المتوسطات . أي أن :

$H_0 : \mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_k$

H_1 : يوجد اختلاف بين متوسطي معالجتين على الأقل

ويلاحظ أن هذه الفروض هي نفس الفروض التي سبق اختبارها في التصميم التام العشوائية . وكذلك تأخذ إحصائية الاختبار نفس الشكل الذي سبق استخدامه في التصميم التام العشوائية :

$$\text{إحصائية الاختبار : ف} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

حيث يحسب البسط (متوسط المربعات بين المعالجات) بنفس الطريقة التي حسب بها في التصميم التام العشوائية في حين يحسب المقام (متوسط مربعات الخطأ) بطريقة مختلفة . حيث يجزئ مجموع المربعات الكلي في تصميم القطاعات الكاملة العشوائية إلى ثلاثة أجزاء بدلا من جزئين . حيث أن :

مجموع المربعات الكلي = مجموع المربعات بين المعالجات + مجموع المربعات بين القطاعات + مجموع مربعات الخطأ

وبالتالي فإن :

مجموع مربعات الخطأ = مجموع المربعات الكلي – مجموع المربعات بين المعالجات – مجموع المربعات بين القطاعات

وهذا يعني أن مجموع مربعات الخطأ في تصميم القطاعات الكاملة العشوائية يساوي مجموع مربعات الخطأ في التصميم التام العشوائية مطروحا

منه مجموع المربعات بين القطاعات . وبناءا على ذلــــك نســتطيع القــول أن تصميم القطاعات الكاملة العشوائية يسمح بإزالة الاختلاف بين القطاعات مــن الاختلاف داخل العينات . وهو من شأنه تقليل متوسط مربعات الخطأ (والــذي يمثل المقام في إحصائية ف) وبالتالي تكون هناك فرصة أكبر لملاحظة الفـوق بين متوسطات المعالجات . ويمكن تلخيـص عنــاصر اختبــار الفــرق بيــن متوسطات المعالجات باستخدام القطاعات الكاملة العشوائية وكذلــك الشــروط الواجب توافرها كالآتي :

اختبار الفرق بين متوسطات ل من المعالجات باستخدام تصميم القطاعات الكاملة العشوائية
الفروض الإحصائية : $H_0 : \mu_1 = \mu_2 = \ldots\ldots = \mu_ل$ H_1 : يوجد اختلاف بين متوسطي معالجتين على الأقل إحصائية الاختبار : ف = متوسط المربعات بين المعالجات / متوسط مربعات الخطأ الشروط : ١ ـ التوزيعات الاحتمالية للمشــاهدات المنــاظرة لكــل توليفــات القطاعات مع المعالجات معتدلة . ٢ ـ تباينات التوزيعات الاحتمالية متساوية . منطقة الرفض : ف > $ف_{(\alpha ، ١٧ ، ٢٧)}$ حيث ١٧ = ل − ١ ، ٢٧ = ن − ط − ل + ١

ويمكن تلخيص الصيغ اللازمة لتحليــل تصميــم القطاعـــات الكاملــة العشوائية في الآتي :

صيغ الحسابات اللازمة لتحليل تصميم القطاعات الكاملة العشوائية

$$\text{معامل التصحيح} = \frac{(\text{مجموع كل المشاهدات})^2}{\text{العدد الكلي للمشاهدات}}$$

$$= \frac{\left(\sum_{و=1}^{ن} س_و\right)^2}{ن}$$

مجموع المربعات الكلي = مجموع مربعات كل المشاهدات – معامل التصحيح

$$= \sum_{و=1}^{ن} س_و^2 - \text{معامل التصحيح}$$

$$\text{مجموع المربعات بين المعالجات} = \left(\begin{array}{c}\text{مجموع مربعات مجاميع المعالجات}\\ \text{مقسوماً كل منها على عدد}\\ \text{المشاهدات للمعالجة}\end{array}\right) - \text{معامل التصحيح}$$

$$= \frac{م_1^2}{ط} + \frac{م_2^2}{ط} + \ldots\ldots\ldots + \frac{م_ل^2}{ط} - \text{معامل التصحيح}$$

$$\text{مجموع المربعات بين القطاعات} = \left(\begin{array}{c}\text{مجموع مربعات مجاميع القطاعات}\\ \text{مقسوما كل منها على عدد}\\ \text{المشاهدات للقطاع}\end{array}\right) - \text{معامل التصحيح}$$

$$= \frac{ق_1^2}{ل} + \frac{ق_2^2}{ل} + \ldots\ldots\ldots + \frac{ق_ط^2}{ل} - \text{معامل التصحيح}$$

مجموع مربعات الخطأ = مجموع المربعات الكلي – مجموع المربعات بين المعالجات – مجموع المربعات بين القطاعات

$$\text{متوسط المربعات بين المعالجات} = \frac{\text{مجموع المربعات بين المعالجات}}{ل - 1}$$

$$\text{متوسط المربعات بين القطاعات} = \frac{\text{مجموع المربعات بين القطاعات}}{\text{ط} - 1}$$

$$\text{متوسط مربعات الخطأ} = \frac{\text{مجموع مربعات الخطأ}}{\text{ن} - \text{ل} - \text{ط} + 1}$$

$$\text{إحصائية الاختبار : ف} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

وجدول (٥ ــ ٦) يوضح تحليل التباين لتصميم القطاعات الكاملة العشوائية :

جدول (٥ ــ ٦)

جدول تحليل التباين لتصميم القطاعات الكاملة العشوائية

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات القطاعات الخطأ	ل – ١ ط – ١ ن – ل – ط + ١	بين المعالجات بين القطاعات الخطأ	بين المعالجات بين القطاعات الخطأ	متوسط مربعات بين المعالجات / متوسط مربعات الخطأ
الكلي	ن – ١	الكلي		

مثال (٤) :

أرادت إحدى الشركات الصناعية الكبرى المنتجة للملابس الجاهزة القيام بتجربة لدراسة تأثير زيادة الأجر / الساعة على إنتاجية العاملين بها فقامت باستخدام أربعة أنظمة للدفع (معالجات)

المعالجة (١) : عدم زيادة الأجر / الساعة .

المعالجة (٢) : زيادة الأجر / الساعة بمقدار ٥٠ قرشا .

المعالجة (٣) : زيادة الأجر / الساعة بمقدار ١٠٠ قرشا .

المعالجة (٤) : زيادة الأجر / الساعة بمقدار ١٥٠ قرشا .

وقامت باختيار أثنى عشر عاملا ، تم تقسيمهم إلى ثلاثة مجموعات على أساس المدة التي قضاها العامل بالشركة على أن يكون بكل مجموعة أربعة عمال توزع عليهم الأربعة معالجات عشوائيا . وقد تم متابعة هؤلاء العمال لمدة ثلاثة أسابيع وقياس إنتاجيتهم بناءا على متوسط عدد الوحدات السليمة المنتجة في الساعة . وجدول (٥ ــ ٧) يلخص النتائج التي تم الحصول عليها . والمطلوب استخدام تحليل التباين لتحديد ما إذا كانت البيانات تدل على وجود فرق بين متوسطات إنتاجية العمال في ظل الأنظمة الأربعة للدفع (استخدم مستوى معنوية ٠,٠٥) .

جدول (٥ ــ ٧)

متوسط إنتاجية العامل في الساعة

المعالجات / المدة التي قضاها العامل بالشركة	(١)	(٢)	(٣)	(٤)	المجموع
أقل من سنة	٢,٤	٣,٠	٣,١	٣,٢	١١,٧
١ ــ ٥ سنوات	٤,٦	٦,١	٥,٩	٥,٧	٢٢,٣
أكثر من خمس سنوات	٥,١	٧,٠	٧,٢	٧,٣	٢٦,٦
المجموع	١٢,١	١٦,١	١٦,٢	١٦,٢	٦٠,٦

الحـــل :

بفرض أن μ_1 ، μ_2 ، μ_3 ، μ_4 تمثل على الترتيب متوســـط إنتاجيـــة العامل في ظل النظام الأول ، الثاني ، الثالث ، الرابع لدفـــع الأجـــر فتكـــون عناصر اختبار تحليل التباين كالآتي :

الفروض الإحصائية :

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1 : يوجد فرق بين متوسطين على الأقل

إحصائية الاختبار : ف = $\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$

الشروط : ١ ــ التوزيعات الاحتمالية لمتوسط عدد الوحدات المنتجة في الساعة المناظرة لكل توليفات المدة التي قضاها العامل فـــي الشـــركة ونظام دفع الأجر معتدلة .

٢ ــ تباينات التوزيعات الاحتمالية متساوية .

منطقة الرفض : ف > ف$_{(\alpha ، v_1 ، v_2)}$

حيث أن : v_1 = ل ــ ١ = ٤ ــ ١ = ٣ ،

v_2 = ن ــ ل ــ ط + ١ = ١٢ ــ ٤ ــ ٣ + ١ = ٦ ، α = ٠,٠٥

∴ ف$_{(\alpha ، v_1 ، v_2)}$ = ف$_{(٠,٠٥ ، ٣ ، ٦)}$ = ٤,٧٦

وباستخدام بيانات الجدول (٥ ــ ٧) نجد أن :

مجـ س$_{ر و}^{٢}$ = (٢,٤)^{٢} + (٣) + ... + (٧,٣)^{٢} = ٣٠٠,٤٢

$$\text{معامل التصحيح} = \frac{(\text{مجـ س}_{و})^{2}}{\text{ن}}$$

$$= \frac{(٦٠,٦)^{2}}{١٢} = ٣٠٦,٠٣$$

مجموع المربعات الكلي = مجـ $\text{س}_{و}^{2}$ ــ معامل التصحيح

$$= ٣٤٠,٤٢ - ٣٠٦,٠٣ = ٣٤,٣٩$$

$$\text{مجموع المربعات بين المعالجات} = \frac{\text{م}_{١}^{2}}{\text{ط}} + \frac{\text{م}_{٢}^{2}}{\text{ط}} + \frac{\text{م}_{٣}^{2}}{\text{ط}} + \frac{\text{م}_{٤}^{2}}{\text{ط}} - \text{معامل التصحيح}$$

$$= \frac{(١٢,١)^{2}}{٣} + \frac{(١٦,١)^{2}}{٣} + \frac{(١٦,٢)^{2}}{٣} + \frac{(١٦,٢)^{2}}{٣} - ٣٠٦,٠٣$$

$$= ٣١٠,١٧ - ٣٠٦,٠٣ = ٤,١٤$$

$$\text{مجموع المربعات بين القطاعات} = \frac{\text{ق}_{١}^{2}}{\text{ل}} + \frac{\text{ق}_{٢}^{2}}{\text{ل}} + \frac{\text{ق}_{٣}^{2}}{\text{ل}} - \text{معامل التصحيح}$$

$$= \frac{(١١,٧)^{2}}{٤} + \frac{(٢٢,٣)^{2}}{٤} + \frac{(٢٦,٦)^{2}}{٤} - ٣٠٦,٠٣$$

$$= ٣٣٥,٤٤ - ٣٠٦,٠٣ = ٢٩,٤١$$

مجموع مربعات الخطأ = مجموع المربعات الكلي ــ مجموع المربعات بين المعالجات ــ مجموع المربعات بين القطاعات

$$= ٣٤,٣٩ - ٤,١٤ - ٢٩,٤١ = ٠,٨٤$$

$$\text{متوسط المربعات بين المعالجات} = \frac{\text{مجموع المربعات بين المعالجات}}{\text{ل} - ١}$$

$$= \frac{٤,١٤}{٤ - ١} = ١,٣٨$$

متوسط المربعات بين القطاعات = $\frac{\text{مجموع المربعات بين القطاعات}}{\text{ط - ١}}$

= $\frac{٢٩,٤١}{٣ - ١}$ = ١٤.٧١

متوسط مربعـــــات الخطـــأ = $\frac{\text{مجموع مربعات الخطأ}}{\text{ن - ل - ط + ١}}$

= $\frac{٠,٨٤}{١٢-٤-٣+١}$ = ٠,١٤

ف = $\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$

= $\frac{١,٣٨}{٠,١٤}$ = ٩,٨٦

وحيث أن قيمة ف المحسوبة (٩,٨٦) أكبر من قيمـــة ف الجدوليـة (٤,٧٦) فأننا نرفض فرض العـــدم بتسـاوي المتوسـطات ونقبـل البديـل بوجود اختلاف بين متوسطين على الأقل وذلك عند مســـتوى معنويـــة ٠,٠٥، وجدول (٥ ــ ٨) يوضح جدول تحليل التباين لمثالنا الحالي :

جدول (٥ ــ ٨)

جدول تحليل التباين لمثال (٤)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٣	٤,١٤	١,٣٨	٩,٨٦
القطاعات	٢	٢٩,٤١	١٤,٧١	
الخطأ	٦	٠,٨٤	٠,١٤	
الكلي	**١١**	**٣٤,٣٩**		

ويمكن إيجاد فترة ثقة للفرق بين متوسطي أي معالجتين مع مراعاة استخدام متوسط مربعات الخطأ كمقدر للتباين σ^2 . أي أن :

$$ع^2 = \text{متوسط مربعات الخطأ}$$

$$= \frac{\text{مجموع مربعات الخطأ}}{ن - ل - ط + ١}$$

وبناءا على ذلك تكون صيغة فترة الثقة كالآتي :

فترة ثقة ١٠٠ (١ − α) ٪ للفرق بين متوسطي معالجتين ($\mu_ر - \mu_ز$)
$(\bar{س}_ر - \bar{س}_ز) \pm ت_{(\frac{\alpha}{2}، ن - ل - ط + ١)}\ ع \sqrt{\frac{١}{ط} + \frac{١}{ط}}$ $= (\bar{س}_ر - \bar{س}_ز) \pm ت_{(\frac{\alpha}{2}، ن - ل - ط + ١)}\ ع \sqrt{\frac{٢}{ط}}$ حيث : $ع = \sqrt{\text{متوسط مربعات الخطأ}}$

مثال (٥) :

في المثال (٤) المطلوب إيجاد فترة ثقة ٩٠٪ للفرق في متوسط الإنتاجية للمعالجات (١) ، (٢)

الحل :

من المثال (٤) نجد أن الوسط الحسابي لإنتاجية عينة العمال الذين لم يتغير أجرهم في الساعة (معالجة (١)) ولعينة العمال الذين زاد أجرهم بمقدار ٥٠ قرشا في الساعة (معالجة (٢)) هما على الترتيب :

$$\bar{س}_١ = \frac{م_١}{ط} = \frac{١٢,١}{٣} = ٤,٠٣$$

$$\bar{س}_٢ = \frac{م_٢}{ط} = \frac{١٦,١}{٣} = ٥,٣٧$$

من جدول تحليل التباين (٥ ــ ٨) نحصل على :

$$ع = \sqrt{\text{متوسط مربعات الخطأ}}$$

$$= \sqrt{٠,١٤} = ٠,٣٧$$

وبالتالي تكون فترة ثقة ٩٠٪ للفرق $\mu_١$ ، $\mu_٢$ على الصورة :

$$(\bar{س}_١ - \bar{س}_٢) \pm ت_{(\frac{\alpha}{٢} ، ن - ل - ط + ١)} ع \sqrt{\frac{٢}{ط}}$$

$$= (\bar{س}_١ - \bar{س}_٢) \pm ت_{(٠,٠٥ ، ٦)} ع \sqrt{\frac{٢}{ط}}$$

$$= (٤,٠٣ - ٥,٣٧) \pm (١,٩٤٣) (٠,٣٧) \sqrt{\frac{٢}{٣}}$$

$$= -١,٣٤ \pm ٠,٥٩$$

$$= (-١,٩٣ ، -٠,٧٥)$$

وهذا يعني أن متوسط إنتاجية العامل بالنسبة للمعالجة (٢) يزيد عن نظيره بالنسبة للمعالجة (١) بمقدار يتراوح بين ٠,٧٥ ، ١,٩٣ .

وبالإضافة إلى ما سبق بشأن إجراء اختبار للفرق بين متوسطات المعالجات فإنه يمكن إجراء اختبار للفرق بين متوسطات القطاعات . فهذا الاختبار سيمكننا من تقرير ما إذا كان تصميم القطاعات قد نجح في تخفيض

حجم الخطأ التجريبي أم لا . بمعنى أنه إذا كان هناك اختلاف بين متوسطات القطاعات فيكون ذلك دليلا على أن الوحدات التجريبية داخل القطاعات أكثر تجانسا منها بين القطاعات مما يبرر الحاجة لاستخدام تصميم القطاعات الكاملة العشوائية . وإجراء الاختبار لمتوسطات القطاعات قريب الشبه جدا من اختبار متوسطات المعالجات . حيث يتم مقارنة الاختلاف بين القطاعات . مقاسا بمتوسط المربعات بين القطاعات ، بالاختلاف الناتج عن الخطأ التجريبي مقاسا بمتوسط مربعات الخطأ . ويمكن تلخيص هذا الاختبار في الآتي :

اختبار الفرق بين متوسطات ط من القطاعات باستخدام تصميم القطاعات الكاملة العشوائية
الفروض الإحصائية : H_0 : متوسطات ط من القطاعات متساوية H_1 : يوجد اختلاف بين متوسطي قطاعين على الأقل إحصائية الاختبار : ف = $\frac{\text{متوسط المربعات بين القطاعات}}{\text{متوسط مربعات الخطأ}}$ الشروط : نفس الشروط السابقة الذكر عند إجراء اختبار الفرق بين متوسطات المعالجات . منطقة الرفض : ف > ف$_{(\alpha ، ح_1 ، ح_2)}$ حيث ح$_1$ = ط − ١ ، ح$_2$ = ن − ل − ط + ١

مثال (٦) :

في المثال (٤) والخاص باستخدام تصميم القطاعات الكاملة العشوائية لمقارنة متوسط إنتاجية العامل في ظل أربعة أنظمة لدفع الأجر . وكانت المدد المختلفة التي قضاها العامل بالشركة ممثلة للقطاعات . اختبر

الفرض بتساوي متوسط إنتاجية العامل في القطاعات الثلاثة . استخدم مستوى معنوية ٠,٠٥ .

الحـــل :

عناصر اختبار الفرق بين متوسطات القطاعات هي :

الفروض الإحصائية :

H_0 : متوسط إنتاجية العامل متساو في القطاعات الثلاثة .

H_1 : يوجد فرق بين متوسطين على الأقل

إحصائية الاختبار : ف = $\frac{\text{متوسط المربعات بين القطاعات}}{\text{متوسط مربعات الخطأ}}$

الشروط : نفس الشروط السابق ذكرها في حل المثال (٤)

منطقة الرفض : ف > ف$_{(\alpha ، \text{ر}_1 ، \text{ر}_2)}$

حيث أن : ر$_1$ = ط ــ ١ = ٣ ــ ١ = ٢ ،

ر$_2$ = ن ــ ل ــ ط + ١ = ٦

∴ ف$_{(\alpha ، \text{ر}_1 ، \text{ر}_2)}$ = ف$_{(٠,٠٥ ، ٢ ، ٦)}$ = ٥,١٤

وفي المثال (٤) تم حساب متوسط المربعات بين القطاعات ، ومتوسط مربعات الخطأ وكانا على الترتيب ١٤,٧١ ، ١٤ وبالتعويض بهذه القيم في إحصائية ف نحصل على :

$$\text{ف} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط مربعات الخطأ}}$$

$$= \frac{١٤,٧١}{٠,١٤} = ١٠٥,٠٧$$

وحيث أن قيمة ف المحسوبة (١٠٥,٠٧) تزيد بدرجة كبيرة عن قيمة ف الجدولية (٥,١٤) فإننا نستنتج أن البيانات كافية لتــأييد الفــرض البديـل بوجود اختلاف في متوسط إنتاجية العامل في القطاعات الثلاثة الممثلــة للمــدة التي قضاها العامل بالشركة . ومن ثم كان القرار باستخدام تصميم القطاعــات الكاملة العشوائية قرارا حكيما . وكان لاستخدام هذا التصميم أثره الواضح فـــي تخفيض مجموع مربعات الخطأ وبالتالي زيادة حجم المعلومات في التجربـــة وجدول (٥ ـــ ٩) يوضح تحليل التباين الكامل لهذه التجربة .

جدول (٥ ـــ ٩)

جدول تحليل التباين الكامل لمثال (٦)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٣	٤,١٤	١,٣٨	٩,٨٦
القطاعات	٢	٢٩,٤١	١٤,٧١	١٠٥,٠٧
الخطأ	٦	٠,٨٤	٠,١٤	
الكلي	١١	٣٤,٣٩		

ونود أن نشير في نهاية هذا البند إلى ضرورة تفســير نتــائج اختبــار الفرق بين متوسطات القطاعات بشيء من الحيطة والحذر وخاصة إذا كـــانت قيمة ف المحسوبة لا تقع في منطقة الرفض . فهذا لا يعني بالضرورة أن تكون

متوسطات القطاعات متساوية وبالتالي لا تكون هناك فائدة من استخدام القطاعات .فالوصول إلى مثل هذا الاستنتاج يناظر القول بقبول فرض العدم والذي يجب تجنبه بسبب عدم معلومية احتمال الوقوع في خطأ من النوع الثاني (قبول فرض العدم وهو غير صحيح في الواقع) وبناءاً على ذلك نستطيع القول أنه في حالة فشل الاختبار في الوصول إلى قرار قاطع بشأن الفروق بين متوسطات القطاعات فيمكن للقائم بالتجربة أن يستخدم تصميم القطاعات الكاملة العشوائية إذا كان لديه اعتقاد بأن الوحدات التجريبية أكثر تجانساً داخل القطاعات منها بين القطاعات .

تمارين (٥)

(١) البيانات التالية لجدول تحليل التباين لتصميم تام العشوائية :

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٦	١٦,٩		
الخطأ				
الكلي	٤١	٤٥,٢		

المطلوب :

أ ــ إكمال جدول تحليل التباين .

ب ــ ما هو عدد المعالجات التي تشملها التجربة .

جـ ــ هل تؤيد هذه البيانات الفرض بوجود اختلاف بين متوسطات المعالجات . استخدم مستوى معنوية α = ٠,٠١ .

د ــ بفرض أن $\overline{س}_١$ = ٣,٧ ، $\overline{س}_٢$ = ٤,١ هل تدل هذه البيانات على وجود اختلاف بين متوسطي المجتمعين $\mu_١$ ، $\mu_٢$ ؟ افترض أنه يوجد ٧ مشاهدات لكل معالجة ، استخدم مستوى معنوية α = ٠,١٠ .

هـ ــ أوجد فترة ثقة ٩٠٪ للفرق بين متوسطي المجتمعين ($\mu_١ - \mu_٢$).

و ــ أوجد فترة ثقة ٩٠٪ لمتوسط المجتمع ($\mu_١$) .

(٢) أراد أحد المكاتب المتخصصة في أعمال مراجعة الدفاتر المالية للمؤسسات الكبرى تقييم الأتعاب التي يتقاضاها المكتب نظير الخدمات التي يقدمها.وكجزء من هذا التقييم أن يقارن تكاليف المراجعة للشركات ذات الأحجام المختلفة ، وقد قرر المكتب قياس حجم المؤسسة العميلة بمقدار مبيعاتها السنوية.وبناء على ذلك تم تقسيم مجتمع العملاء إلى ثلاثة مجتمعات فرعية :

س = مجتمع العملاء الذين تزيد مبيعاتهم عن ٢٥٠ مليون جنيه .

ص = مجتمع العملاء الذين تتراوح مبيعاتهم بين ١٠٠ مليون ، ٢٥٠ مليون جنيه .

ع = مجتمع العملاء الذين تقل مبيعاتهم عن ١٠٠ مليون جنيه

وقد قام المكتب باختيار عينة عشوائية مكونة من عشرة عملاء من كل مجتمع من هذه المجتمعات و الجدول التالي يلخص تكاليف المراجعة (بآلاف الجنيهات) :

تكاليف المراجعة (بآلاف الجنيهات)		
س	ص	ع
٢٥٠	١٠٠	٨٠
١٥٠	١٥٠	١٢٥
٢٧٥	٧٥	٢٠
١٠٠	٢٠٠	١٨٦
٤٧٥	٥٥	٥٢
٦٠٠	٨٠	٩٢
١٥٠	١١٠	٨٨
٨٠٠	١٦٠	١٤١
٣٢٥	١٣٢	٧٦
٢٣٠	٢٣٣	٢٠٠

المطلوب :

أ ـ استخدام تحليل التباين لتحديد ما إذا كان هناك فرق معنوي بين متوسطات تكاليف المراجعة للمجتمعات الثلاثة من العملاء ، استخدم مستوى معنوية α = ٠,٠٥ .

ب ـ إيجاد فترة ثقة ٩٥٪ للفرق (مب ـ مص) .

جـ ـ أذكر الشروط اللازم توافرها حتى يمكن عمل الاستدلال الإحصائي في الجزئين أ ، ب .

(٣) أراد مدير أحد الشركات التي تقوم بالاستعانة بعدد كبير من مندوبي المبيعات دراسة تأثير نظام دفع الأجر على كمية المبيعات التي يحققها المندوب في الشهر.حيث يوجد ثلاثة أنظمة لدفع الأجر (العمولة ، الأجر الثابت ، أجر ثابت منخفض + عمولة) ولإجراء هذه الدراسة قام المدير بسحب عينة عشوائية من مندوبي المبيعات الذين يطبق عليهم الأنظمة المختلفة لدفع الأجر ، ويلخص الجدول التالي المبيعات (بالجنيهات) التي حققها هؤلاء المندوبين خلال الشهر :

عمولة	أجر ثابت	أجر ثابت + عمولة
٤٢٥	٤٢٠	٤٣٠
٥٠٧	٤٤٨	٤٩٢
٤٥٠	٤٣٧	٤٧٠
٤٩٣	٤٣٢	٥٠١
٤٦٦	٤٤٤	
٤٩٢		

المطلوب :

أ ـ هل تؤيد هذه البيانات وجود اختلاف بين متوسط مبيعات المندوبين في ظل الأنظمة المختلفة لدفع الأجر .

ب ـ أوجد فترة ثقة ٩٠٪ لمتوسط مبيعات المندوبين الذين يتقاضون (اجر ثابت + عمولة) .

جـ ـ ايجاد فترة ثقة ٩٠٪ للفرق في متوسط المبيعات للمندوبين الذين يتقاضون (أجر ثابت + عمولة) وهؤلاء الذين يتقاضون أجر ثابت .

(٤) الجدول التالي لتحليل التباين لتصميم قطاعات كاملة العشوائية :

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
المعالجات	٣	٢٨,٢		
القطاعات	٥		١٣,٨	
الخطأ		٣٤,١		
الكلي				

المطلوب :

أ ـ أكمل جدول تحليل التباين .

ب ـ هل تدل هذه البيانات على وجود اختلاف بين متوسطات المعالجات .

جـ ـ هل تؤيد هذه البيانات ضرورة استخدام تصميم القطاعات لهذه التجربة .

د ـ إذا كانت الأوساط الحسابية للعينات ف ، ق هي $\bar{س}_{ف}$ = ٩,٧ ، $\bar{س}_{ق}$ = ١٢,١ على الترتيب ، أوجد فترة ثقة ٩٠ ٪ للفرق ($\mu_{ف} - \mu_{ق}$) .

الفصل السادس

الاستدلال الإحصائي باستخدام أسلوب كا^2

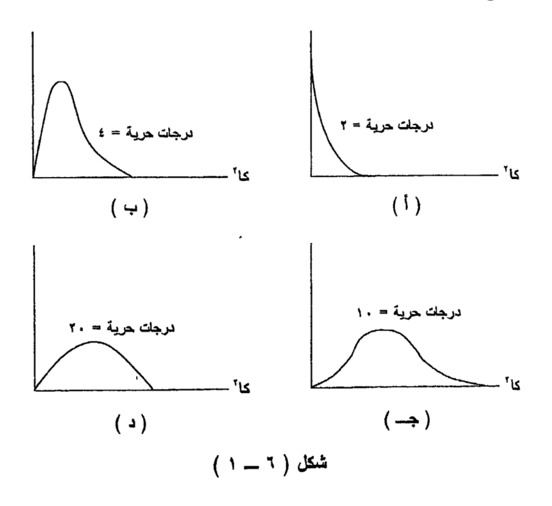
(٦ ــ ١) مقدمة :

استعرضنا في الفصول السابقة طرق عمل الاستدلال الإحصائي عن بعض معالم المجتمع بالاعتماد على توزيع z . أو ت أو ف . ولكن في بعض الأحيان قد لا نستطيع الاعتماد على مثل هذه التوزيعات . ونلجأ إلى ما يعرف بأسلوب كاي تربيع Chi-Square Technique . والذي يعتمد على توزيع كاي تربيع .

ويعتبر توزيع كاي تربيع (كا^2) من التوزيعات المستمرة الهامة . ولإيضاح فكرة هذا التوزيع باختصار : نفترض أن لدينا متغيراً عشوائياً (س) له توزيع معتدل وسطه الحسابي يساوي μ وانحرافه المعياري يساوي σ . فإذا ما طرحنا من هذا المتغير وسطه الحسابي وقسمنا على انحرافه المعياري نحصل على متغير (z) يتبع للتوزيع المعتدل المعياري بوسط حسابي يساوي الصفر وانحراف معياري يساوي الواحد الصحيح . وبتربيع المتغير z نحصل على متغير يتبع لتوزيع مستمر شديد الالتواء تجاه اليمين وتتراوح قيم هذا المتغير بين صفر و ما لانهاية . ويُعرف هذا التوزيع بتوزيع كا^2 بدرجات حرية تساوي الواحد الصحيح . فإذا ما كان لدينا ن من المتغيرات المستقلة z_1 ، z_2 ، . . . . ، $z_ن$ فإن مجموع مربعات هذه المتغيرات يتبع لتوزيع كا^2 بدرجات حرية ن . ومتوسط هذا التوزيع يساوي عدد درجات الحرية ، وتباينه يساوي ضعف عدد درجات الحرية . ويوضح الشكل (٦ ــ ١) توزيع كا^2 عند

درجات حرية مختلفة . ويلاحظ أن التواء المنحنى تجاه اليمين يقل بزيادة عـــدد درجات الحرية ويقترب هذا التوزيع من الاعتدال إذا كانت درجات الحرية أكبر من أو تساوي ٣٠ وتوجد جداول لتوزيع كا^{٢} تعطي قيم كا^{٢} المناظرة للمساحات المختلفة للطرف الأيمن ولدرجات الحرية المختلفة .

وسوف نتناول في هذا الفصل بعض التطبيقات علـــى اســـتخدام كـــا^{٢} وهي : الاستدلال الإحصائي عن أكثر من نسبتي مجتمعين ، اختبار الاســـتقلال بين متغيرين وصفيين ، اختبار جودة التوفيق ، الاستدلال الإحصائي عن تبـاين المجتمع .



شكل (٦ ــ ١)

٦ ــ ٢) الاستدلال الإحصائي عن أكثر من نسبتي مجتمعين :

إذا فرضنا أن أحد الشركات التجارية أرادت القيام بدراسة تفضيلات المستهلك لعلامات زيوت الطعام المختلفة . وكان لدى هذه الشركة ثلاثة أنواع من الزيوت (أ ، ب ، جـ) . فقامت باختيار عينة عشوائية مكونة من ١٥٠ مستهلكاً . وباستجواب كل مستهلك عن النوع الذي يفضله ، أمكن الحصول على النتائج الموضحة بالجدول (٦ ــ ١) . هل تؤيد هذه البيانات تفضيل المستهلك لنوع معين ؟

جدول (٦ ــ ١)

نوع الزيت	أ	ب	جـ	المجموع
عدد المستهلكين	٦١	٥٣	٣٦	١٥٠

للإجابة على هذا السؤال يجب التعرف أولاً على التوزيع الاحتمالي لهذه النوعية من البيانات،حيث أنها تتبع لتوزيع يُعرف بالتوزيع الاحتمالي المتعدد الحدود Multinomial وتتلخص خواصه في الآتي :

خواص التوزيع الاحتمالي المتعدد الحدود :
١ ــ عدد المحاولات التي تتكون منها التجربة يساوي ن ٢ ــ عدد النواتج الممكنة في كل محاولة يساوي ك ٣ ــ احتمالات النواتج في كل محاولة ثابتة (أي لا تتغير من محاولة إلى أخرى) وتساوي θ_1 ، θ_2 ، . . . . ، $\theta_{ك}$. حيث $\theta_1 + \theta_2 + ... + \theta_{ك} = ١$ ٤ ــ المحاولات مستقلة .

ويلاحظ أن خواص التوزيع المتعدد الحدود قريبة الشبه من خواص توزيع ذي الحدين . ونستطيع القول أن توزيع ذي الحدين هو حالة خاصة من التوزيع المتعدد الحدود (حيث ك = ٢) .

وغالباً ما تكون القيم الحقيقية للاحتمالات θ_1 ، θ_2 . . . . ، $\theta_{ك}$ مجهولة وبالتالي يكون هدفنا الأساسي هو عمل استدلال إحصائي عن هذه الاحتمالات . ففي مثالنا الحالي والخاص بتفضيل المستهلك والذي يحقق شروط التوزيع المتعدد الحدود إذا فرضنا أن :

θ_1 = نسبة المستهلكين الذين يفضلون النوع أ .

θ_2 = نسبة المستهلكين الذين يفضلون النوع ب .

θ_3 = نسبة المستهلكين الذين يفضلون النوع جـ .

فيكون هدفنا هو اختبار فرض العدم بأنه لا يوجد تفضيل لأي نوع من الأنواع الثلاثة في مقابل البديل بأنه يوجد تفضيل لنوع أو أكثر من هذه الأنواع . وبالتالي يمكن صياغة هذه الفروض على الصورة الآتية :

$H_0 : \theta_1 = \theta_2 = \theta_3 = \frac{1}{3}$ (لا يوجد تفضيل) .

H_1 : يوجد نسبة واحدة على الأقل تزيد عن $\frac{1}{3}$ (يوجد تفضيل) .

فإذا كان فرض العدم صحيحاً فإننا نتوقع أن يكون عدد المستهلكين لكل نوع يمثل $\frac{1}{3}$ حجم العينة تقريباً . بمعنى أنه إذا كانت $ن_1$ ، $ن_2$ ، $ن_3$ ترمز على التوالي إلى عدد المستهلكين الذين يفضلون النوع أ ، النوع ب ، النوع جـ والتي سنطلق عليها التكرارات المشاهدة فإن التكرارات المتوقعة المناظرة لها يمكن إيجادها كالآتي :

$$توقع\ (ن_1) = ن\ \theta_1$$

$$= ١٥٠ \left(\frac{1}{3} \right) = ٥٠$$

، كذلك أيضاً : توقع ($ن_٢$) = توقع ($ن_٣$) = ٥٠

وإحصائية الاختبار في هذه الحالة والتي تقيس درجة الاختلاف بين بيانات العينة (التكرارات المشاهدة) والبيانات تحت شرط صحة فرض العدم (التكرارات المتوقعة) هي :

$$كا^٢ = \frac{[ن_١ - توقع (ن_١)]^٢}{توقع (ن_١)} + \frac{[ن_٢ - توقع (ن_٢)]^٢}{توقع (ن_٢)} + \frac{[ن_٣ - توقع (ن_٣)]^٢}{توقع (ن_٣)}$$

$$= \frac{(ن_١ - ٥٠)^٢}{٥٠} + \frac{(ن_٢ - ٥٠)^٢}{٥٠} + \frac{(ن_٣ - ٥٠)^٢}{٥٠}$$

ويلاحظ أنه كلما بعُدت التكرارات المشاهدة عن المتوقعة ، كلما حصلنا على قيمة أكبر لإحصائية الاختبار $كا^٢$. وهذا يعني أن القيمة الكبيرة لإحصائية الاختبار تعطي مؤشراً على عدم صحة فرض العدم .

وحتى نتمكن من أخذ القرار برفض أو قبول فرض العدم لابد أن تكون على علم بشكل التوزيع العيني لإحصائية الاختبار $كا^٢$. وقد أمكن إثبات أنه تحت شرط صحة فرض العدم فإن التوزيع العيني لإحصائية الاختبار يقترب من توزيع $كا^٢$ بدرجات حرية ٧ = ٣ – ١ = ٢ . وعلى فرض أننا نريد إجراء الاختبار عند مستوى معنوية α = ٠,٠٥ فإننا سنرفض فرض العدم H_0 إذا كان :

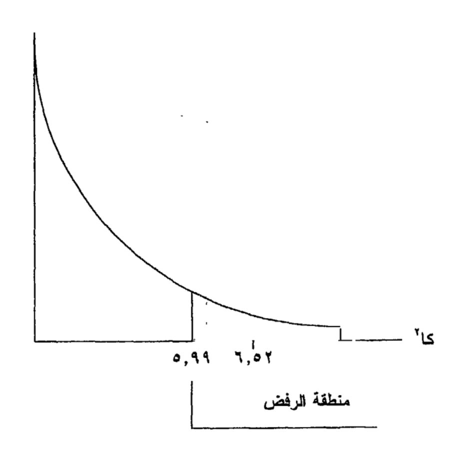
$$كا^٢ > كا^٢_{(٠,٠٥ ، ٢)} = ٥,٩٩ \quad \text{(من جدول } كا^٢ \text{)}$$

وباستخدام بيانات عينة المستهلكين التي تم سحبها عشوائياً نحصل على :

$$كا^٢ = \frac{(ن_١ - ٥٠)^٢}{٥٠} + \frac{(ن_٢ - ٥٠)^٢}{٥٠} + \frac{(ن_٣ - ٥٠)^٢}{٥٠}$$

$$= \frac{(٦١ - ٥٠)^٢}{٥٠} + \frac{(٥٣ - ٥٠)^٢}{٥٠} + \frac{(٣٦ - ٥٠)^٢}{٥٠} = ٦,٥٢$$

وحيث أن كا^{2} المحسوبة (٦,٥٢) تقع في منطقة الرفض (كما هو موضح بالشكل (٦ ــ ٢) فإننا نستنتج رفض فرض العدم وقبول الفرض البديل بتفضيل المستهلك لنوع أو أكثر من أنواع الزيوت .



شكل (٦ ــ ٢)

ويمكن تلخيص عناصر اختبار الفرض الإحصائي المتعلق بأكثر من نسبتي مجتمعين في الآتي :

اختبار الفرض الإحصائي المتعلق بأكثر من نسبتي مجتمعين

الفروض الإحصائية :

H_0 : $\theta_1 = \theta_{1o}$ ، $\theta_2 = \theta_{2o}$ ، .... ، $\theta_k = \theta_{ko}$

(حيث θ_{1o} ، θ_{2o} ، ...... ، θ_{ko} هي القيم المفترضة لنسب المجتمعات)

H_1 : يوجد على الأقل نسبة واحدة لا تساوي القيمة المفترضة

إحصائية الاختبار : $كا^2 = مجـ \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$

منطقة الرفض : $كا^2 > كا^2_{(\alpha ، v)}$ (حيث v = ك − ١)

الشروط :

١ ــ أن يكون توزيع المجتمعات الأصلية معتدلاً .

٢ ــ أن يكون حجم العينة كبيراً بدرجة تكفي لجعل التكرار المتوقع مساوياً للقيمة ٥ على الأقل .

ويلاحظ أنه في حالة عدم توافر الشرط ٢ فأننا ندمج التكرارات المتوقعة وبالتالي التكرارات المشاهدة المناظرة لها وذلك حتى نستوفي ذلك الشرط . ومن الطبيعي أن يحسب عدد درجات الحرية بعد هذا الإدماج .

مثال (١) :

إذا كانت اللائحة الداخلية لإحدى الشركات تنص على نظام معين لتوزيع الحوافز السنوية على موظفي الشركة . حيث يعتمد هذا النظام على الدرجات التي تمنح للموظف بواسطة رئيسة المباشر في العمل.فالموظف الذي تزيد درجته عن ٨٠ يستحق الحد الأقصى للحوافز السنوية ، والذي تتراوح درجته بين ٥٠ ، ٨٠ يستحق الحوافز العادية ، الذي تقل درجته عن ٥٠ لا يستحق أي حوافز . وكانت الشركة قد وضعت في خطتها أن يستحق

الحد الأقصى ٢٥٪ من الموظفين ، يستحق الحوافز العادية ٦٥٪ . بينما لا تستحق النسبة الباقية (١٠٪) من الموظفين أي حوافز.وبعد سنة واحدة من استخدام هذا النظام تم سحب عينة عشوائية من ٦٠٠ موظف بالشركة . وكان توزيع الحوافز السنوية كما هو موضح بالجدول (٦ ــ ٢) . هل تدل هذه البيانات على وجود اختلاف معنوي بين توزيع الحوافز وفقا للائحة الداخلية للشركة وتوزيعها وفقا لما وضعته الشركة في خطتها . استخدم مستوى معنوية α = ٠,٠١ .

جدول (٦ ــ ٢)

لا يستحق الحوافز	يستحق حوافز عادية	يستحق الحد الأقصى
٤٢	٣٦٥	١٩٣

الحـــل :

بفرض أن :

θ_1 = نسبة الموظفين الذين لا يستحقون أي حوافز .

θ_2 = نسبة الموظفين الذين يستحقون حوافز عادية .

θ_3 = نسبة الموظفين الذين يستحقون الحد الأقصى للحوافز .

فبالتالي يكون فرض العدم والذي يمثل خطة الشركة في توزيع الحوافز على الصورة :

H_0 : θ_1 = ٠,١٠ ، θ_2 = ٠,٦٥ ، θ_3 = ٠,٢٥

ويكون الفرض البديل على الصورة :

H_1 : يوجد نسبة واحدة على الأقل لا تتفق مع خطة الشركة .

إحصائية الاختبار : $كا^{2} = مجـ \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^{2}}{\text{التكرار المتوقع}}$

منطقة الرفض : $كا^{2} > كا^{2}_{(v,\alpha)} = كا^{2}_{(٢،٠,٠١)} = ٩,٢١$

وباستخدام بيانات العينة يمكن حساب إحصائية الاختبار كما هو موضح بالجدول (٦ ــ ٣)

جدول (٦ ــ ٣)

التكرار المشاهد	التكرار المتوقع	مشاهد – متوقع	(مشاهد – متوقع)^{2}	$\frac{(\text{مشاهد} - \text{متوقع})^{2}}{\text{متوقع}}$
٤٢	٦٠٠ (٠,١٠) = ٦٠	١٨-	٣٢٤	٥,٤
٣٦٥	٦٠٠ (٠,٦٥) = ٣٩٠	٢٥-	٦٢٥	١,٦
١٩٣	٦٠٠ (٠,٢٥) = ١٥٠	٤٣	١٨٤٩	١٢,٣
٦٠٠	٦٠٠	صفر		$كا^{2}$ = ١٩,٣

وحيث أن قيمة $كا^{2}$ المحسوبة تقع في منطقة الرفض فأننا نستنتج رفض فرض العدم وقبول البديل بأن توزيع الحوافز السنوية لا يتفق مع خطة الشركة. وذلك عند مستوى معنوية α = ٠,٠١

وبالإضافة إلي إجراء الاختبار عن نسب المجتمعات فإنه يمكن إيجاد فترة ثقة لأي نسبة من هذه النسب . فعلى سبيل المثال يمكن إيجاد فترة ثقة ٩٥٪ لنسبة الموظفين الذين يستحقون الحد الأقصى للحوافز كالآتي :

$$\hat{\theta}_{٣} \pm ١,٩٦ \, \sigma_{\hat{\theta}_{٣}}$$

$$\approx \hat{\theta}_{٣} \pm ١,٩٦ \sqrt{\frac{\hat{\theta}_{٣}(١ - \hat{\theta}_{٣})}{ن_{٣}}}$$

(حيث $\hat{\theta}_{٢} = \frac{ن_{٢}}{ن} = \frac{١٩٣}{٦٠٠} = ٠,٣٢$)

$$= ٠,٣٢ \pm ١,٩٦ \sqrt{\frac{٠,٣٢ (١ - ٠,٣٢)}{٦٠٠}}$$

$$= ٠,٣٢ \pm ٠,٠٤$$

$$= (٠,٢٨ ، ٠,٣٦)$$

وهذا يعني أن تتراوح نسبة الموظفين الذين يستحقون الحد الأقصى للحوافز فيما بين ٢٨٪ ، ٣٦٪ ومن ذلك يتضح أنه يجب زيادة عدد الدرجات اللازمة للحصول على الحد الأقصى للحوافز حتى يمكن تحقيق النسبة ٢٥٪ والتي وضعتها الشركة في خطتها .

(٦ ــ ٣) اختبار الاستقلال بين متغيرين وصفيين :

إذا فرضنا أن أحد الباحثين الاقتصاديين أراد دراسة العلاقة بين حجم السيارة المشتراة حديثا والشركة المنتجة لها . فقام باختيار عينة عشوائية من ١٠٠٠ مشتري . وتم تصنيفهم وفقا لحجم السيارة المشتراة والشركة المنتجة لها . ويوضح الجدول (٦ ــ ٤) البيانات التي تم الحصول عليها .

جدول (٦ ــ ٤)

الشركة المنتجة للسيارة / حجم السيارة	أ	ب	جـ	د	المجموع
صغير	١٥٧	٦٥	١٨١	١٠	٤١٣
متوسط	١٢٦	٨٢	١٤٢	٤٦	٣٩٦
كبير	٥٨	٤٥	٦٠	٢٨	١٩١
المجموع	٣٤١	١٩٢	٣٨٣	٨٤	١٠٠٠

والآن بفرض أن جدول (٦ ــ ٥) يمثل احتمال وقوع كل ناتج من نواتج الجدول (٦ ــ ٤) . حيث تشير θ_{11} إلى احتمال شراء سيارة صغيرة الحجم ومنتجة بواسطة الشركة أ ، θ_{21} تشير إلى احتمال شراء سيارة صغيرة الحجم ومنتجة بواسطة الشركة ب وهكذا بالنسبة لبقية خلايا الجدول . كما يشير مجموع الاحتمالات في كل صف أو كل عمود إلى ما يعرف بالاحتمالات الهامشية Marginal Probabilities . حيث θ_{1} تمثل احتمال شراء سيارة صغيرة الحجم ، $\theta_{أ}$ تمثل احتمال شراء سيارة منتجة بواسطة الشركة أ . وهكذا بالنسبة لبقية الاحتمالات الهامشية .

جدول (٦ ــ ٥)

حجم السيارة \ الشركة المنتجة للسيارة	أ	ب	جـ	د	المجموع
صغير	θ_{11}	θ_{21}	θ_{31}	θ_{41}	θ_{1}
متوسط	θ_{12}	θ_{22}	θ_{32}	θ_{42}	θ_{2}
كبير	θ_{13}	θ_{23}	θ_{33}	θ_{43}	θ_{3}
المجموع	$\theta_{أ}$	$\theta_{ب}$	$\theta_{جـ}$	$\theta_{د}$	١

ولتحديد ما إذا كان هناك علاقة بين المتغيرين (حجم السيارة والشركة المنتجة لها) فإننا نرجع إلى تعريف الاستقلال بين المتغيرين في حالة الجدول المزدوج . حيث يقال أن المتغيرين مستقلان إذا كان الاحتمال في أي خلية يساوي حاصل ضرب الاحتمالات الهامشية المناظرة . وبناءاً على ذلك فإننا نستطيع القول أنه إذا كان حجم السيارة مستقل عن الشركة المنتجة فسنجد أن :

$\theta_{١١} = \theta_{١} \times \theta_{ا}$ ، $\theta_{٢١} = \theta_{١} \times \theta_{ب}$

$\theta_{٣١} = \theta_{١} \times \theta_{جـ}$ ، $\theta_{٤١} = \theta_{١} \times \theta_{د}$

وهكذا بالنسبة لبقية الاحتمالات في كل خلايا الجدول .

ولإجراء الاختبار الإحصائي لفرض الاستقلال بين المتغيرين نستخدم نفس الفكرة التي استخدمت في البند السابق (٦ ـ ٢) والتي تعتمد على إيجاد التكرار المتوقع لكل خلية بافتراض صحة هذا الفرض . حيث أن :

توقع (ن ١١) = التكرار المتوقع في الخلية التي تقع بالصف الأول والعمود الأول

$= ن \, \theta_{١١}$

$= ن \times \theta_{١} \times \theta_{ا}$

$= ن \left(\frac{ن_{١}}{ن} \right) \left(\frac{ن_{ا}}{ن} \right)$ { حيث $ن_{١}$ = مجموع الصف الأول ، $ن_{ا}$ = مجموع العمود الأول }

$= \frac{ن_{١} \times ن_{ا}}{ن}$

وبنفس الأسلوب نحصل على :

توقع ($ن_{٢١}$) $= \frac{ن_{١} \times ن_{ب}}{ن}$ { حيث $ن_{ب}$ = مجموع العمود الثاني }

توقع ($ن_{٣١}$) $= \frac{ن_{١} \times ن_{جـ}}{ن}$ { حيث $ن_{جـ}$ = مجموع العمود الثالث }

. .
. .
. .
. .

توقع ($ن_{٤٣}$) $= \frac{ن_{٣} \times ن_{د}}{ن}$ { حيث $ن_{٣}$ = مجموع الصف الثالث ، $ن_{د}$ = مجموع العمود الرابع }

وباستخدام بيانات الجدول (٦ ـ ٤) نحصل على :

$$\text{توقع}\ (ن_{١١}) = \frac{٣٤١ \times ٤١٣}{١٠٠٠} = ١٤٠,٨٣٣$$

$$\text{توقع}\ (ن_{٢١}) = \frac{١٩٢ \times ٤١٣}{١٠٠٠} = ٧٩,٢٩٦$$

$$\text{توقع}\ (ن_{٣١}) = \frac{٣٨٣ \times ٤١٣}{١٠٠٠} = ١٥٨,١٧٩$$

.
.
.
.

$$\text{توقع}\ (ن_{٤٣}) = \frac{٨٤ \times ١٩١}{١٠٠٠} = ١٦,٠٤٤$$

ويوضح جدول (٦ ـ ٦) التكرارات المشاهدة والمتوقعة (بداخل قوسين) لمثالنا الحالي :

جدول (٦ ـ ٦)

الشركة المنتجة للسيارة / حجم السيارة	أ	ب	جـ	د
صغير	١٥٧ (١٤٠,٨٣٣)	٦٥ (٧٩,٢٩٦)	١٨١ (١٥٨,١٧٩)	١٠ (٣٤,٦٩٢)
متوسط	١٢٦ (١٣٥,٠٣٦)	٨٢ (٧٦,٠٣٢)	١٤٢ (١٥١,٦٦٨)	٤٦ (٣٣,٢٦٤)
كبير	٥٨ (٦٥,١٣١)	٤٥ (٣٦,٦٧٢)	٦٠ (٧٣,١٥٣)	٢٨ (١٦,٠٤٤)

وبعد إيجاد التكرارات المتوقعة نستخدم إحصائية كا^2 للمقارنة بين التكرار المشاهد والمتوقع في كل من خلايا الجدول كالآتي :

$$\text{كا}^2 = \text{مجـ} \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

$$= \frac{(١٥٧ - ١٤٠,٨٣٣)^2}{١٤٠,٨٣٣} + \frac{(٦٥ - ٧٩,٢٩٦)^2}{٧٩,٢٩٦} + \dots + \frac{(٢٨ - ١٦,٠٤٤)^2}{١٦,٠٤٤}$$

$$= ٤٥,٨١$$

حيث تشير القيم الكبيرة لإحصائية كا^2 إلى وجود اختلاف كبير بين التكرارات المشاهدة والمتوقعة وهذا من شأنه التشكيك في صحة الفرض باستقلال المتغيرين . وللحكم على قيمة الإحصائية كا^2 فإنه يتم مقارنتها بقيمة كا^2 الجدولية بدرجات حرية مساوية (عدد الصفوف – ١) (عدد الأعمدة – ١) . ففي مثالنا الحالي تكون درجات الحرية = (٣ – ١) (٤ – ١) = ٦ . وبالتالي فإنه عند مستوى معنوية α = ٠,٠٥ ، نرفض الفرض باستقلال المتغيرين إذا كان :

$$\text{كا}^2 > \text{كا}^2_{(٠,٠٥ ، ٦)} = ١٢,٥٩$$

وحيث أن قيمة كا^2 المحسوبة (٤٥,٨١) أكبر من قيمة كا^2 الجدولية (١٢,٥٩) فإننا نرفض فرض العدم باستقلال المتغيرين . وهذا يعني أنه توجد علاقة بين حجم السيارة المشتراة حديثا والشركة المنتجة لها .

ومما سبق نستطيع تلخيص لختبار الاستقلال بين متغيرين وصفين في الآتي :

اختبار الاستقلال بين متغيرين وصفيين

الفروض الإحصائية :

H_0 : لا توجد علاقة بين المتغيرين (المتغيران مستقلان)

H_1 : توجد علاقة بين المتغيرين (المتغيران غير مستقلين)

إحصائية الاختبار :

$$كا^2 = مجـ \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

$$\text{حيث التكرار المتوقع} = \frac{\text{مجموع الصف} \times \text{مجموع العمود}}{\text{المجموع الكلي}}$$

منطقة الرفض : $كا^2 > كا^2_{(v, \alpha)}$

حيثv = (عدد الصفوف – ١) (عدد الأعمدة – ١)

الشرط : يجب أن يكون حجم العينة كبيرا بدرجة تكفي لجعل التكرار المتوقع في كل خلية لا يقل عن ٥ وفي حالة عدم توافر هذا الشرط فإننا نلجأ إلى إدماج التكرارات المتوقعة وبالتالي التكرارات المشاهدة المناظرة لها . وذلك حتى نستوفي هذا الشرط وتحسب درجات الحرية بعد هذا الإدماج .

مثال (٢) :

أراد قسم المراقبة على جودة الإنتاج في أحد الشركات أن يحدد إذ كان هناك علاقة بين سنوات الخبرة ومعدل الخطأ والذي يقاس بعدد الوحدات المعيبة في كل ١٠٠٠ وحدة منتجة بواسطة العامل . فقام باختيار عينة عشوائية مكونة من مائة عامل ويوضح جدول (٦ ــ ٧) نتائج هذه التجربة . فهل تؤيد هذه النتائج وجود علاقة بين سنوات الخبرة . ومعدل الخطأ في إنتاج العامل ؟ استخدم مستوى معنوية α = ٠,٠٥ .

جدول (٦ ــ ٧)

سنوات الخبرة / معدل الخطأ في الإنتاج	أقل من سنة	سنة إلى أقل من ٥	٥ ــ ١٠	المجموع
مرتفع	٦	٩	٩	٢٤
متوسط	٩	١٩	٢٣	٥١
منخفض	٧	٨	١٠	٢٥
المجموع	٢٢	٣٦	٤٢	١٠٠

الحـــل :

نبدأ أولا بحساب التكرارات المتوقعة في كل خلية بـــافتراض صحـــة فرض العدم بأن المتغيرين مستقلان . أي أن :

توقع ($ن_{١١}$) = $\frac{\text{مجموع الصف الأول} \times \text{مجموع العمود الأول}}{\text{المجموع الكلي}}$

= $\frac{٢٤ \times ٢٢}{١٠٠}$ = ٥,٢٨

توقع ($ن_{٢١}$) = $\frac{\text{مجموع الصف الأول} \times \text{مجموع العمود الثاني}}{\text{المجموع الكلي}}$

= $\frac{٢٤ \times ٣٦}{١٠٠}$ = ٨,٦٤

وهكذا بالنسبة لجميع خلايـا الجـدول . يوضـح جـدول (٦ ــ ٨) التكرارات المشاهدة والمتوقعة (بداخل قوسين) .

جدول (٦ ــ ٨)

معدل الخطأ \ سنوات الخبرة	أقل من سنة	سنة إلى أقل من ٥	٥ – ١٠
مرتفع	٦ (٥,٢٨)	٩ (٨,٦٤)	٩ (١٠,٠٨)
متوسط	٩ (١١,٢٢)	١٩ (١٨,٣٦)	٢٣ (٢١,٤٢)
منخفض	٧ (٥,٥٠)	٨ (٩)	١٠ (١٠,٥٠)

والخطوة التالية لذلك هي إجراء اختبار الاستقلال بين المتغيرين :

الفروض الإحصائية :

H_0 : لا توجد علاقة بين سنوات الخبرة ومعدل الخطأ .

H_1 : توجد علاقة بين سنوات الخبرة ومعدل الخطأ .

إحصائية الاختبار :

$$\text{كا}^2 = \text{مجـ} \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$\text{كا}^2 > \text{كا}^2_{(\alpha ، v)} = \text{كا}^2_{(٠,٠٥ ، ٤)} = ٩,٤٩$$

حيث v = (عدد الصفوف – ١) (عدد الأعمدة – ١)

= (٣ – ١) (٣ – ١) = ٤

وباستخدام بيانات الجدول (٦ ــ ٨) يمكن حساب إحصائية الاختبار كا^2 كما هو موضح بالجدول (٦ ــ ٩) .

جدول (٦ ــ ٩)

التكرار المشاهد	التكرار المتوقع	مشاهد ــ متوقع	(مشاهد ــ متوقع)٢	(مشاهد ــ متوقع)٢ / متوقع
٦	٥,٢٨	٠,٧٢	٠,٥١٨٤	٠,٠٩٨
٩	٨,٦٤	٠,٣٦	٠,١٢٩٦	٠,٠١٥
٩	١٠,٠٨	١,٠٨-	١,١٦٦٤	٠,١١٦
٩	١١,٢٢	٢,٢٢-	٤,٩٢٨٤	٠,٤٣٩
١٩	١٨,٣٦	٠,٦٤	٠,٤٠٩٦	٠,٠٢٢
٢٣	٢١,٤٢	١,٥٨	٢,٤٩٦٤	٠,١١٧
٧	٥,٥٠	١,٥٠	٢,٢٥٠٠	٠,٤٠٩
٨	٩,٠٠	١,٠٠-	١,٠٠٠٠	٠,١١١
١٠	١٠,٥٠	٠,٥٠-	٠,٢٥٠٠	٠,٠٢٤
١٠٠	١٠٠	صفر		كا٢ = ١,٣٥١

وحيث أن قيمة كا٢ المحسوبة (١,٣٥١) أقل من قيمة كا٢ الجدولية (٩,٤٩) فإننا نستنتج أن بيانات العينة غير كافية لتأييد الفرض البديل بأنه توجد علاقة بين سنوات الخبرة ومعدل الخطأ في إنتاج العامل وذلك عند مستوى معنوية α = ٠,٠٥ .

(٦ ــ ٤) اختبار جودة التوفيق : Goodness of Fit test

يستخدم اختبار جودة التوفيق لتحديد التوزيع الاحتمالي الذي تتبع له بيانات المجتمع أو لاختبار مدى تبعية البيانات لتوزيع معين . ونورد فيما يلي بعض الأمثلة التطبيقية على استخدام هذا الاختبار .

(٦ ــ ٤ ــ ١) اختبار جودة التوفيق لتوزيع ذي الحدين :

مثال (٣) :

أرادت إحدى شركات الاستيراد والتصدير دراسة توزيع الوحدات التالفة في شحنة من سلعة معينة وكانت معبأة في صناديق حيث يحتوي الصندوق الواحد على ٦ وحدات من السلعة . وقد أدعى مصدر هذه السلعة أن الوحدات التالفة في هذه الشحنة تتبع لتوزيع ذي الحدين ولاختبار صحة هذا الادعاء تم فحص الوحدات التالفة في عينة عشوائية مكونة من ٢٠٠ صندوق ، وتم الحصول على النتائج الموضحة بالجدول (٦ ــ ١٠) المطلوب اختبار ادعاء مصدر هذه السلعة . استخدم مستوى معنوية ٠,٠٥ .

جدول (٦ ــ ١٠)

عدد الوحدات التالفة في الصندوق	صفر	١	٢	٣	٤	٥	٦	المجموع
عدد الصناديق	٣١	٥١	٧٠	٣٢	٩	٥	٢	٢٠٠

الحــــل :

حيث أننا نرغب في اختبار صحة إدعاء المصدر بأن الوحدات التالفة في الشحنة تتبع لتوزيع ذي الحدين فيجب أولا تحديد معالم هذا التوزيع وهي عدد المحاولات (ن) ، واحتمال النجاح (θ) . بالنسبة إلى ن فنجد أنها تمثل أقصى رقم لعدد الوحدات التالفة في الصندوق وهو ٦ وبالنسبة لاحتمال النجاح (θ) فيمكن تقديره بإيجاد نسبة التالف في العينة . أي أن :

$$\hat{\theta} = \frac{\text{عدد الوحدات التالفة في العينة}}{\text{العدد الكلي للوحدات في العينة}}$$

$$= \frac{\text{صفر} \times ٣١ + ١ \times ٥١ + ٢ \times ٧٠ + ٣ \times ٣٢ + ٤ \times ٩ + ٥ \times ٥ + ٦ \times ٢}{٢٠٠ \times ٦}$$

= ٠,٣

ومن ذلك نستطيع إيجاد التكرارات المتوقعة بافتراض صحة فرض العدم بأن عدد الوحدات التالفة يتبع توزيع ذي الحدين كما هو موضح بجدول (٦ ـ ١١) .

جدول (٦ ـ ١١)

عدد الوحدات التالفة في الصندوق (س)	ح(س = س) = ^{ن}ق$_{س}$ $\theta^{س}$ $(١-\theta)^{ن-س}$ حيث ن=٦ ، θ=٠,٣	التكرار المتوقع ٢٠٠ × ح (س = س)
صفر	ح (س = صفر) = ^{٦}ق$_{صفر}$ (٠,٣)^{صفر}(٠,٧)^{٦}=٠,١١٧٦	٢٠٠×٠,١١٧٦ = ٢٣,٥٢
١	ح (س = ١) = ^{٦}ق$_{١}$ (٠,٣)^{١}(٠,٧)^{٥}=٠,٣٠٢٥	٢٠٠×٠,٣٠٢٥ = ٦٠,٥٠
٢	ح (س = ٢) = ^{٦}ق$_{٢}$ (٠,٣)^{٢}(٠,٧)^{٤}=٠,٣٢٤١	٢٠٠×٠,٣٢٤١ = ٦٤,٨٢
٣	ح (س = ٣) = ^{٦}ق$_{٣}$ (٠,٣)^{٣}(٠,٧)^{٣}=٠,١٨٥٢	٢٠٠×٠,١٨٥٢ = ٣٧,٠٤
٤	ح (س = ٤) = ^{٦}ق$_{٤}$ (٠,٣)^{٤}(٠,٧)^{٢}=٠,٠٥٩٥	٢٠٠×٠,٠٥٩٥ = ١١,٩٠
٥	ح (س = ٥) = ^{٦}ق$_{٥}$ (٠,٣)^{٥}(٠,٧)^{١}=٠,٠١٠٢	٢٠٠×٠,٠١٠٢ = ٢,٠٤] ١٤,٠٨
٦	ح (س = ٦) = ^{٦}ق$_{٦}$ (٠,٣)^{٦}(٠,٧)^{صفر}=٠,٠٠٠٧	٢٠٠×٠,٠٠٠٧ = ٠,١٤

يلاحظ أنا دمجنا التكرارات المتوقعة عند س = ٤ ، ٥ ، ٦ حتى نستوفي شرط تطبيق كا^{٢} وهو ألا يقل التكرار المتوقع في أي خلية عن العدد ٥ .

والخطوة التالية لإيجاد التكرارات المتوقعة هي إجراء اختبار الفرض بأن البيانات تتبع لتوزيع ذي الحدين .

الفروض الإحصائية :

H_0 : عدد الوحدات التالفة في الشحنة يتبع لتوزيع ذي الحدين .

H_1 : عدد الوحدات التالفة لا يتبع لتوزيع ذي الحدين .

إحصائية الاختبار :

$$كا^2 = مج \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$كا^2 > كا^2_{(\nu ، \alpha)} = كا^2_{(٠,٠٥ ، ٣)} = ٧,٨١٥$$

[حيث ν = عدد التكرارات المتوقعة – ١ – عدد المعالم التي تم تقديرها
= ٥ – ١ – ١ = ٣]

وباستخدام بيانات الجدول (٦ ــ ١١) يمكن حساب إحصائية الاختبار كما هو موضح بالجدول (٦ ــ ١٢) .

جدول (٦ ــ ١٢)

عدد الوحدت التالف في الصندوق	التكرار المشاهد	التكرار المتوقع	مشاهد – متوقع	(مشاهد – متوقع)^2	(مشاهد – متوقع)^2 / متوقع
صفر	٣١	٢٣,٥٢	٧,٤٨	٥٥,٥٩٠	٢,٣٧٩
١	٥١	٦٠,٥٠	٩,٥٠-	٩٠,٢٥٠	١,٤٩٢
٢	٧٠	٦٤,٨٢	٥,١٨	٢٦,٨٣٢	٠,٤١٤
٣	٣٢	٣٧,٠٤	٥,٠٤-	٢٥,٤٠٢	٠,٦٨٦
٤ – ٦	١٦	١٤,٠٨	١,٩٢	٣,٦٨٦	٠,٢٦٢
	٢٠٠	٢٠٠	صفر		كا^2 = ٥,٢٣٣

وحيث أن قيمة كا^2 المحسوبة (٥,٢٣٣) أقل من قيمة كا^2 الجدولية (٧,٨١٥) فإننا نستنتج أن البيانات تتبع لتوزيع ذي الحدين .

(٦ ــ ٤ ــ٢) اختبار جودة التوفيق لتوزيع بواسون :

مثال (٤) :

في إحدى الدراسات التي أجريت لمعرفة التوزيـــع الاحتمـــالي لعـــدد المرضى الذين يصلون خلال ساعة إلى العيادة الخارجية بأحد المستشفيات تـــم اختيار عينة عشوائية مكونة من ٥٠ ساعة عمل ، وتم تسجيل عدد المرضـــى الذين يصلون خلال الساعة . والجدول (٦ ــ ١٣) يلخـــص النتـــائج التـــي أمكن الحصول عليها . المطلوب اختبار فرض العـــدم بـــأن عـــدد المرضـــى الذين يصلون خلال الساعة يتبع توزيع بواســـون بوســـط حســـابي λ = ٣,٨ . استخدم معنوية ٠,٠١ .

جدول (٦ ــ ١٣)

عدد المرضى الذين يصلون خلال الساعة	صفر	١	٢	٣	٤	٥	٦	٧	٨ فأكثر	المجموع
عدد الساعات	صفر	١	٥	٨	١٥	٩	٧	٣	٢	٥٠

الحـــل :

نبدأ أولا بإيجاد التكرارات المتوقعة بافتراض صحة فرض العدم بـــأن عدد المرضى الذين يصلون خلال الساعة يتبع لتوزيع بواسون بوسط حســـابي λ = ٣,٨ (كما هو موضح بجدول (٦ ــ ١٤)) .

جدول (٦ ـ ١٤)

عدد المرضى الذين يصلون خلال الساعة (س)	ح (سـ = س) = هـ^(-λ) λ^(س) / |س ، حيث λ = ٣,٨	التكرار المتوقع = ٥٠×ح(سـ = س)	
صفر	ح (سـ = صفر) = هـ^(-٣,٨) (٣,٨)^(صفر) / |صفر = ٠,٠٢٢٤	١,١٢=٠,٠٢٢٤×٥٠	٥,٣٧
١	ح (سـ = ١) = هـ^(-٣,٨) (٣,٨)^(١) / |١ = ٠,٠٨٥٠	٤,٢٥=٠,٠٨٥٠×٥٠	
٢	ح (سـ = ٢) = هـ^(-٣,٨) (٣,٨)^(٢) / |٢ = ٠,١٦١٥	٨,٠٧٥=٠,١٦١٥×٥٠	
٣	ح (سـ = ٣) = هـ^(-٣,٨) (٣,٨)^(٣) / |٣ = ٠,٢٠٤٦	١٠,٢٣=٠,٢٠٤٦×٥٠	
٤	ح (سـ = ٤) = هـ^(-٣,٨) (٣,٨)^(٤) / |٤ = ٠,١٩٤٤	٩,٧٢=٠,١٩٤٤×٥٠	
٥	ح (سـ = ٥) = هـ^(-٣,٨) (٣,٨)^(٥) / |٥ = ٠,١٤٧٧	٧,٣٨٥=٠,١٤٧٧×٥٠	
٦	ح (سـ = ٦) = هـ^(-٣,٨) (٣,٨)^(٦) / |٦ = ٠,٠٩٣٦	٤,٦٨=٠,٠٩٣٦×٥٠	٩,٢٣
٧	ح (سـ = ٧) = هـ^(-٣,٨) (٣,٨)^(٧) / |٧ = ٠,٠٥٠٨	٢,٥٤=٠,٠٥٠٨×٥٠	
٨ فأكثر	ح (سـ ≥ ٨) = ١ − ح (سـ < ٨) = ١ − ٠,٩٦ = ٠,٠٤٠٠	٢,٠٠= ٠,٠٤٠٠×٥٠	

والخطوة التالية اختيار الفرض بأن البيانات تتبع لتوزيع بواسون بوسط حسابي λ = ٣,٨ :

الفروض الإحصائية :

H_0 : عدد المرضى الذين يصلون خلال الساعة يتبع لتوزيع بواسون بوسط حسابي = ٣,٨

H_1 : عدد المرضى الذين يصلون خلال الساعة لا يتبع لتوزيع بواسون بوسط حسابي = ٣,٨

إحصائية الاختبار :

$$\text{كا}^2 = \text{مج} \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$\text{كا}^2 > \text{كا}^2_{(\alpha \text{،} v)} = \text{كا}^2_{(0{,}01 \text{،} 5)} = 15{,}086$$

{ حيث v = ٦ − ١ = ٥ }

وباستخدام بيانات الجدول (٦ ـــ ١٤) يمكن حساب إحصائية الاختبار $\text{ك}_{٠}^{2}$ كما هو موضح بالجدول (٦ ـــ ١٥) .

جدول (٦ ــ ١٥)

عدد المرضى الذين يصلون خلال ساعة	التكرار المشاهد	التكرار المتوقع	مشاهد – متوقع	(مشاهد – متوقع)^2	(مشاهد – متوقع)^2 / متوقع
صفر – ١	١	٥,٣٧	٤,٣٧٠-	١٩,٠٩٧	٣,٥٥٦
٢	٥	٨,٠٧٥	٣,٠٧٥-	٩,٤٥٦	١,١٧١
٣	٨	١٠,٢٣	٢,٢٣٠-	٤,٩٧٣	٠,٤٨٦
٤	١٥	٩,٧٢	٥,٢٨	٢٧,٨٧٨	٢,٨٦٨
٥	٩	٧,٣٨٥	١,٦١٥	٢,٦٠٨	٠,٣٥٣
٦ فأكثر	١٢	٩,٢٢	٢,٧٨	٧,٧٢٨	٠,٨٣٨
المجموع	٥٠	٥٠	صفر		كا^2 = ٩,٢٧٢

وحيث أن قيمة كا^2 المحسوبة (٩,٢٧٢) أقل من قيمة كا^2 الجدولية (١٥,٠٨٦) فإننا نستنتج أن عدد المرضى الذين يصلون خلال الساعة إلى العيادة الخارجية بالمستشفى يتبع لتوزيع بواسون بوسط حسابي λ = ٣,٨ .

(٦ ــ ٤ ــ٣) اختبار جودة التوفيق للتوزيع المعتدل :

مثال (٥) :

أراد أحد الباحثين الاقتصاديين دراسة التوزيع الاحتمالي للدخل السنوي لطائفة المحامين . فقام باختيار عينة عشوائية مكونة من ٩٠ محام ، ويلخص جدول (٦ ــ ١٦) النتائج التي تم الحصول عليها . المطلوب اختبار الفرض بأن الدخل السنوي لطائفة المحامين يتبع لتوزيع معتدل بوسط حسابي μ = ٥٠ ألف جنيهاً ، انحراف معياري σ = ١٠ آلاف جنيهاً . استخدم مستوى معنوية ٠,٠٥ .

جدول (٦ ــ ١٦)

فئات الدخل الثانوي (بآلاف الجنيهات)	أقل من ١٠	-١٠	-٢٠	-٣٠	-٤٠	-٥٠	-٦٠	-٧٠	٨٠ فأكثر	المجموع
عدد المحامين	٥	٩	١٥	٢٣	٢٠	٨	٦	٣	١	٩٠

الحـــــل :

نبدأ أولا بإيجاد التكرارات المتوقعة بافتراض صحة فرض العدم بـــأن الدخل السنوي يتبع لتوزيع معتدل بوسط حسابي μ = ٥٠ ألف جنيها ، انحراف معياري σ = ١٠ آلاف جنيها (كما هو موضح بجدول (٦ ــ ١٧)) .

جدول (٦ ــ ١٧)

فئات الدخل السنوي (بآلاف الجنيهات)	الاحتمال (المساحة تحت المنحنى المعتدل المعياري) حيث Z = $\frac{س - \mu}{\sigma}$ = $\frac{س - ٥٠}{١٠}$	التكرار المتوقع = ٩٠×ح(س)	
أقل من ١٠	ح (س < ١٠) = ح (Z < -٤) = صفر	٩٠×صفر=صفر	
-١٠	ح(١٠ < س< ٢٠)= ح(-٤ < Z < -٣)= ٠,٠٠١٤	٩٠×٠,٠٠١٤=٠,١٢٦	
-٢٠	ح(٢٠ < س< ٣٠)= ح(-٣ < Z < -٢)= ٠,٠٢١٤	٩٠×٠,٠٢١٤=١,٩٢٦	١٤,٢٨٣
-٣٠	ح(٣٠ < س< ٤٠)= ح(-٢ < Z < -١)= ٠,١٣٥٩	٩٠×٠,١٣٥٩=١٢,٢٣١	
-٤٠	ح(٤٠ < س< ٥٠)= ح(-١ < Z < صفر)= ٠,٣٤١٣	٩٠×٠,٣٤١٣=٣٠,٧١٧	
-٥٠	ح(٥٠ < س< ٦٠)= ح(صفر < Z < ١)= ٠,٣٤١٣	٩٠×٠,٣٤١٣=٣٠,٧١٧	
-٦٠	ح(٦٠ < س< ٧٠)= ح(١ < Z < ٢)= ٠,١٣٥٩	٩٠×٠,١٣٥٩=١٢,٢٣١	
-٧٠	ح(٧٠ < س< ٨٠)= ح(٢ < Z < ٣)= ٠,٠٢١٤	٩٠×٠,٠٢١٤=١,٩٢٦	١٤,٢٨٣
٨٠ فأكثر	ح(س > ٨٠)= ح(Z > ٣)= ٠,٠٠١٤	٩٠×٠,٠٠١٤=٠,١٢٦	

والخطوة التالية هي اختبار الفرض بتبعية البيانات للتوزيـــع المعتــدل بوسط حسابي μ = ٥٠ ألف جنيها ، انحراف معياري σ = ١٠ آلاف جنيها :

الفروض الإحصائية :

H_0 : الدخل السنوي لطائفة المحامين يتبع لتوزيع معتدل بوسط حسابي μ = ٥٠ ألف جنيها ، انحراف معياري σ = ١٠ آلاف جنيها

H_1 : الدخل السنوي لا يتبع لتوزيع معتدل بوسط حسابي μ = ٥٠ ألف جنيها ، انحراف معياري σ = ١٠ آلاف جنيها

إحصائية الاختبار :

$$\text{كا}^2 = \text{مجـ} \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض :

$$\text{كا}^2 > \text{كا}^2_{(\alpha \text{،} v)} = \text{كا}^2_{(0.05 \text{،} 3)} = 7.815$$

{ حيث v = ٤ – ١ = ٣ }

وباستخدام بيانات الجدول (٦ ــ ١٧) يمكن حساب إحصائية الاختبار كا^٢ كما هو موضح بالجدول (٦ ــ ١٨) .

جدول (٦ ــ ١٨)

فئات الدخل	التكرار المشاهد	التكرار المتوقع	مشاهد – متوقع	(مشاهد – متوقع)٢	(مشاهد – متوقع)٢ / متوقع
أقل من ٤٠	٥٢	١٤,٢٨٣	٣٧,٧١٧	١٤٢٢,٥٧٢	٩٩,٥٩٩
٤٠ –	٢٠	٣٠,٧١٧	١٠,٧١٧-	١١٤,٨٥٤	٣,٧٣٩
٥٠ –	٨	٣٠,٧١٧	٢٢,٧١٧-	٥١٦,٠٦٢	١٦,٨٠١
٦٠ فأكثر	١٠	١٤,٢٨٣	٤,٢٨٣-	١٨,٣٤٤	١,٢٨٤
المجموع	٩٠	٩٠	صفر		كا٢ = ١٢١,٤٢٣

وحيث أن قيمة كا٢ المحسوبة (١٢١,٤٢٣) أكــبر مــن قيمــة كــا٢ الجدولية (٧,٨١٥) فإننا نرفض فرض العدم . وهذا لا يعني بالضرورة عـــدم تبعية الدخل السنوي لطائفة المحامين للتوزيع المعتدل بل من الممكن أن يتبـــع الدخل السنوي للتوزيع المعتدل ولكن بوسط حسابي يختلف عن ٥٠ ألف جنيــها وانحراف معياري يختلف عن ١٠ آلاف جنيها .

(٦ ــ ٤ ــ٤) اختبار جودة التوفيق للتوزيع المنتظم :

مثال (٦) :

قام أحد المسئولين عن مراكز المطافئ بــأحد المــدن بتســجيل آخــر ٧٠٠ حريق نشب في المدينة وكان توزيعهم على مدار أيام الأسبوع كما هـــو موضح بالجدول (٦ ــ ١٩) . المطلوب اختبار فرض العـــدم بـــأن حـــدوث الحرائق في المدينة يتوزع توزيعا منتظما على مدار أيام الأســبوع . اســتخدم مستوى معنوية ٠,٠١ .

جدول (٦ ــ ١٩)

أيام الأسبوع	السبت	الأحد	الإثنين	الثلاثاء	الأربعاء	الخميس	الجمعة	المجموع
عدد الحرائق	٨٥	١٢٩	٧٢	٩١	١٢٣	١٣٥	٦٥	٧٠٠

الحـــل :

الفروض الإحصائية :

H_0 : تتوزع الحرائق توزيعا منتظما على مدار أيام الأسبوع .

H_1 : لا تتوزع الحرائق توزيعا منتظما على مدار أيام الأسبوع .

$$\text{كا}^2 = \text{مجـ} \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^2}{\text{التكرار المتوقع}}$$

منطقة الرفض : $\text{كا}^2 > \text{كا}^2_{(\alpha ، \nu)} = \text{كا}^2_{(٠,٠١ ، ٦)} = ١٦,٨١٢$

ولحساب إحصائية الاختبار يجب أولا حساب التكرار المتوقع بافتراض صحة فرض العدم بأن عدد الحرائق يتوزع بالتساوي على أيام الأسبوع كما هو موضح بجدول (٦ ــ ٢٠) .

جدول (٦ ــ ٢٠)

أيام الأسبوع	التكرار المشاهد	التكرار المتوقع	مشاهد – متوقع	(مشاهد – متوقع)^2	(مشاهد – متوقع)^2 / متوقع
السبت	٨٥	١٠٠	١٥-	٢٢٥	٢,٢٥
الأحد	١٢٩	١٠٠	٢٩	٨٤١	٨,٤١
الإثنين	٧٢	١٠٠	٢٨-	٧٨٤	٧,٨٤
الثلاثاء	٩١	١٠٠	٩-	٨١	٠,٨١
الأربعاء	١٢٣	١٠٠	٢٣	٥٢٩	٥,٢٩
الخميس	١٣٥	١٠٠	٣٥	١٢٢٥	١٢,٢٥
الجمعة	٦٥	١٠٠	٣٥-	١٢٢٥	١٢,٢٥
المجموع	٧٠٠	٧٠٠	صفر		كا^2 = ٤٩,١٠

وحيث أن قيمة $كا^2$ المحسوبة (٤٩,١٠) أكبر من قيمة $كا^2$ الجدوليـــة (١٦,٨١٢) فإننا نرفض فرض العدم ونقبل البديل بأن الحرائـــق لا تتـــوزع بانتظام على مدار أيام الأسبوع وذلك عند مستوى معنوية ٠,٠١ .

وبالإضافة إلى ما سبق من أمثلة على اختبار جودة التوفيـــق لبعـض التوزيعات المعروفة فإنه من الممكن استخدام هذا الاختبار أيضـــا لأي توزيـــع آخر لا يتبع إلى مثل هذه التوزيعات المعروفة . والمثال التالي يوضح ذلك .

مثال (٧) :

أرادت إدارة برامج المنوعات بالتليفزيون مقارنة أنماط عادات المشاهدة لبرامج التليفزيون في عام ١٩٨٠ بما كـــانت عليـــه عـــام ١٩٧٠ م . فقـــامت باختيار عينة عشوائية من ١٠٠ مشاهد في عام ١٩٧٠ ، عام ١٩٨٠ وجـــدول (٦ ــ ٢١) يلخص النتائج التي تم الحصول عليها ، فهل تريد هذه البيانـــــات اختلاف أنماط عادات المشاهدة في عام ١٩٨٠ عنها في عام ١٩٧٠ ، اســـتخدم مستوى معنوية ٠,٠٥ .

جدول (٦ ــ ٢١)

عدد ساعات مشاهدة التليفزيون في الأسبوع	عدد المشاهدين	
	١٩٧٠	١٩٨٠
صفر ــ	٣	٩
٣ ــ	٤	٣
٥ ــ	١٠	٩
٧ ــ	٢٠	٢١
١٥ ــ	٤٥	٤٤
٢٥ فأكثر	١٨	١٤
المجموع	١٠٠	١٠٠

الحـــل :

الفروض الإحصائية :

H_0 : لا تختلف أنماط عادات المشاهدة في عـــام ١٩٨٠ عنـــها فـــي عام ١٩٧٠ .

H_1 : تختلف أنماط عادات المشـــاهدة فـــي عـــام ١٩٨٠ عنـــها فـــي عام ١٩٧٠ .

$$كا^{٢} = مجـ \frac{(\text{التكرار المشاهد} - \text{التكرار المتوقع})^{٢}}{\text{التكرار المتوقع}}$$

وبافتراض صحة فرض العدم تكون التكرارات المتوقعـــة هـــي عـــدد المشاهدين في عام ١٩٧٠ .

منطقة الرفض :

$$كا^{٢} > كا^{٢}_{(\alpha ، ٧)} = كا^{٢}_{(٠٫٠٥ ، ٤)} = ٩٫٤٨٨$$

{ حيث ٧ = ٥ – ١ = ٤ لأننا دمجنا الفئتين الأولى والثانية حتـــى لا يقل التكرار المتوقع عن ٥ } .

وباستخدام بيانات العينة يمكن حساب إحصائية الاختبار كا٢ كما هـــو موضح بجدول (٦ ــ ٢٢) .

جدول (٦ ــ ٢٢)

عدد ساعات مشاهدة التليفزيون في الأسبوع	التكرار المشاهد	التكرار المتوقع	مشاهد – متوقع	(مشاهد – متوقع)^2	$\frac{(مشاهد – متوقع)^2}{متوقع}$
صفر –	١٢	٧	٥	٢٥	٣,٥٧١
٥ –	٩	١٠	١–	١	٠,١٠٠
٧ –	٢١	٢٠	١	١	٠,٠٥٠
١٥ –	٤٤	٤٥	١–	١	٠,٠٢٢
٢٥ فأكثر	١٤	١٨	٤–	١٦	٠,٨٨٩
المجموع	١٠٠	١٠٠	صفر		كا^2 = ٤,٦٣٢

وحيث أن كـــا^2 المحســوبة (٤,٦٣٢) أقــل مــن كــا^2 الجدوليـــة (٩,٤٨٨) فإننا نستنتج أن أنماط عادات المشاهدة في عام ١٩٨٠ لم تختلـــف عنها في عام ١٩٧٠ .

(٤ ــ ٥) الاستدلال الإحصائي عن تباين المجتمع σ^2 :

من المعروف أن ع^2 = مجـ (س – $\overline{س}$)^2 / ن – ١ يستخدم كمقدر غير متحيز لتباين المجتمع σ^2 . ولكن لكي نحدد درجة دقة في التقدير لابـــد مـــن معرفة شكل التوزيع العيني للمقدر ع^2 . وقد تم إثبات أن الكمية :

$$\frac{ع^2 (ن - ١)}{\sigma^2}$$ تتبع لتوزيع كا^2 بدرجات حرية (ن – ١)

ومن ذلك يمكن اشتقاق فترة ثقة ١٠٠ (١ – α) ٪ لتباين المجتمع كالآتي :

$$\text{ح}\left(\text{كا}^2_{(1-\frac{\alpha}{2}\,،\,\text{ن}-1)} < \text{كا}^2 < \text{كا}^2_{(\frac{\alpha}{2}\,،\,\text{ن}-1)}\right) = 1-\alpha$$

$$= \text{ح}\left(\text{كا}^2_{(1-\frac{\alpha}{2}\,،\,\text{ن}-1)} < \frac{\text{ع}^2\,(\text{ن}-1)}{\sigma^2} < \text{كا}^2_{(\frac{\alpha}{2}\,،\,\text{ن}-1)}\right) = 1-\alpha$$

$$= \text{ح}\left(\frac{1}{\text{كا}^2_{(\frac{\alpha}{2}\,،\,\text{ن}-1)}} < \frac{\sigma^2}{\text{ع}^2\,(\text{ن}-1)} < \frac{1}{\text{كا}^2_{(1-\frac{\alpha}{2}\,،\,\text{ن}-1)}}\right) = 1-\alpha$$

$$= \text{ح}\left(\frac{\text{ع}^2\,(\text{ن}-1)}{\text{كا}^2_{(\frac{\alpha}{2}\,،\,\text{ن}-1)}} < \sigma^2 < \frac{\text{ع}^2\,(\text{ن}-1)}{\text{كا}^2_{(1-\frac{\alpha}{2}\,،\,\text{ن}-1)}}\right) = 1-\alpha$$

فترة ثقة ١٠٠ (١ − α) ٪ لتباين المجتمع σ^2

$$\frac{\text{ع}^2\,(\text{ن}-1)}{\text{كا}^2_{(\frac{\alpha}{2}\,،\,\text{ن}-1)}} < \sigma^2 < \frac{\text{ع}^2\,(\text{ن}-1)}{\text{كا}^2_{(1-\frac{\alpha}{2}\,،\,\text{ن}-1)}}$$

مثال (٨) :

أراد مدير أحد شركات صناعة الأدوية معرفة تباين وزن قرص الدواء الجديد . فقام باختيار عينة عشوائية مكونة من ٣٠ قرص . وحسب تباين وزن القرص فكان ٣ ميللجرام . فعلى فرض أن أوزان هذه الأقراص تتبع لتوزيع معتدل تقريبا فأوجد فترة ثقة ٩٠ ٪ لتباين وزن قرص الدواء .

الحـــل :

فترة ثقة ١٠٠ (١ − α) ٪ لتباين المجتمع σ^2 هي :

$$\frac{\text{ع}^2\,(\text{ن}-1)}{\text{كا}^2_{(\frac{\alpha}{2}\,،\,\text{ن}-1)}} < \sigma^2 < \frac{\text{ع}^2\,(\text{ن}-1)}{\text{كا}^2_{(1-\frac{\alpha}{2}\,،\,\text{ن}-1)}}$$

وحيث أن درجات الحرية = ن − ١ = ٣٠ − ١ = ٢٩ ، α = ٠,١٠

فباستخدام جداول كا² نحصل على :

$كا^٢_{(\frac{\alpha}{٢} ، ن - ١)} = كا^٢_{(٠,٠٥ ، ٢٩)} = ٤٢,٥٥٧$

$كا^٢_{(١ - \frac{\alpha}{٢} ، ن - ١)} = كا^٢_{(٠,٩٥ ، ٢٩)} = ١٧,٧٠٨$

وبالتالي فإن فترة ثقة ٩٠ ٪ لتباين وزن قرص الدواء تكـــون علـــى الصورة :

$$\frac{٣ (٢٩)}{٤٢,٥٥٧} < \sigma^٢ < \frac{٣ (٢٩)}{١٧,٧٠٨}$$

$$٢,٠٤ < \sigma^٢ < ٤,٩١$$

وهذا يعني أننا نثق بدرجة ٩٠ ٪ في أن يقع تباين وزن قرص الدواء بين ٢,٠٤ ، ٤,٩١ ميللجرام .

وبالإضافة إلى إيجاد فترة ثقة لتباين المجتمع فإنـــه يمكـــن إجـــراء اختبارات الفروض الإحصائية عن تباين المجتمع .

اختبار الفرض الإحصائي عن تباين المجتمع ($\sigma^٢$)	
اختبار ذو طرف واحد الفروض الإحصائية : $H_0 : \sigma^٢ = \sigma_0^٢$ $H_1 : \sigma^٢ < \sigma_0^٢$ (أو $H_1 : \sigma^٢ > \sigma_0^٢$) حيث $\sigma_0^٢$ القيمة المفترضة لتباين المجتمع إحصائية الاختبار : $كا^٢ = \frac{ع^٢ (ن - ١)}{\sigma_0^٢}$ منطقة الرفض : $كا^٢ < كا^٢_{(١ - \alpha ، ن-١)}$ أو $كا^٢ > كا^٢_{(\alpha ، ن-١)}$ إذا كان $H_1 : \sigma^٢ > \sigma_0^٢$	اختبار ذو طرفين الفروض الإحصائية : $H_0 : \sigma^٢ = \sigma_0^٢$ $H_1 : \sigma^٢ \neq \sigma_0^٢$ إحصائية الاختبار : $كا^٢ = \frac{ع^٢ (ن - ١)}{\sigma_0^٢}$ منطقة الرفض : $كا^٢ < كا^٢_{(١- \frac{\alpha}{٢} ، ن-١)}$ أو $كا^٢ > كا^٢_{(\frac{\alpha}{٢} ، ن-١)}$
الشرط : أن يكون توزيع المجتمع المسحوب منه العينة معتدلا (أو قريبا من الاعتدال) .	

مثال (٩) :

أراد مدير أحد البنوك أن يطبق سياسة الصف الواحد في تقديم الخدمة للعملاء حيث يدخل العميل في الصف بمجرد وصوله للبنك وبعد ذلك يتم توزيع العميل على الشباك المختص بتقديم الخدمة . وقد وجد أنه بالرغم من أن هذه السياسة لن تؤثر على متوسط وقت انتظار العميل في البنك إلا أن المدير يفضل هذه السياسة لأنها تقلل من تباين وقت الانتظار . وكان يتوقع أن يكون هذا التباين أقل من تباين وقت الانتظار في حالة استخدام سياسة الصفوف المتعددة (حيث كان التباين = ٦٤) وللتأكد من ذلك قام بتطبيق سياسة الصف الواحد على عينة عشوائية مكونة من ٣٠ عميل وحسب الانحراف المعياري لوقت انتظار العميل فكان مساويا ٣ دقائق . فهل تؤيد بيانات العينة اعتقاد المدير . استخدم مستوى معنوية ٠,٠٥ .

الحـــل :

حيث أننا نرغب في تحديد ما إذا كان تباين وقت انتظار العميل في البنك يقل عن ٦٤ دقيقة فإن نوع الاختبار الذي يجب استخدامه هو اختبار ذو طرف أيسر وتكون عناصره كالآتي :

الفروض الإحصائية :

$H_0 : \sigma^٢ = ٦٤$

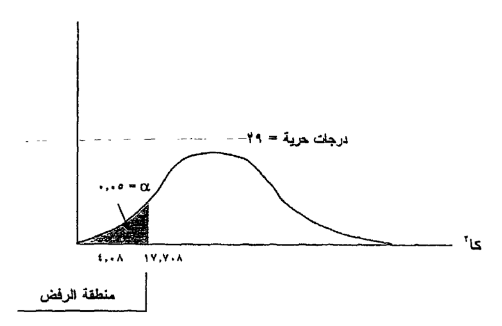
$H_1 : \sigma^٢ < ٦٤$

إحصائية الاختبار :

$$كا^٢ = \frac{ع^٢ \,(ن - ١)}{\sigma^٢_0}$$

منطقة الرفض كا^{2} < كا^{2}$_{(1-\alpha ، ن-1)}$ = كا^{2}$_{(0.95 ، 29)}$ = ١٧,١٠٨

(كما هو موضح بالشكل (٦ ــ ٣))



شكل (٦ ــ ٣)

وباستخدام بيانات العينة نحصل على :

$$\text{كا}^{2} = \frac{\text{ع}^{2}(\text{ن}-1)}{\sigma_{0}^{2}}$$

$$= \frac{(9)(29)}{64} = 4.08$$

وحيث أن قيمة كا^{2} المحسوبة (٤,٠٨) تقع في منطقة الرفض (كما هو واضح في الشكل (٦ ــ ٣)) فإننا نقبل الفرض البديل بأن σ^{2} < ٦٤ وبالتالي يكون اعتقاد المدير صحيحا في أن سياسة الصف الواحد سوف تؤدي إلى تقليل تباين وقت انتظار العميل .

مثال (١٠) :

أعلنت إحدى شركات التشييد والمباني أن تباين قوة ضغط الخرسانة المسلحة للمباني التي تقوم بإنشائها يزيد عن ٨٠ كيلو جرام لكل متر مربع وتم لختبار عينة عشوائية مكونة من ٢٥ عجينة للخرسانة وحسب الانحراف المعياري لقوة الضغط فكان ٨ كيلو جرام لكل متر مربع . فهل تؤيد هذه البيانات إدعاء الشركة . استخدم مستوى معنوية ٠,٠٥ .

الحل :

الفروض الإحصائية :

$H_0 : \sigma^2 = ٨٠$

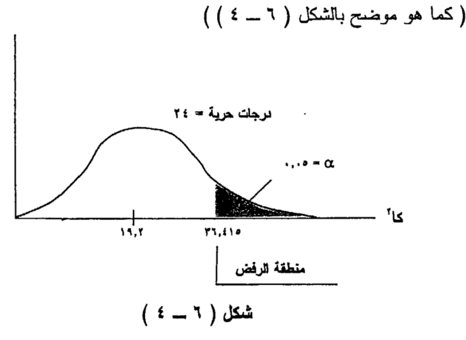
$H_1 : \sigma^2 > ٨٠$

إحصائية الاختبار :

$$كا^2 = \frac{ع^2 \,(ن - ١)}{\sigma_0^2}$$

منطقة الرفض $كا^2 > كا^2_{(\alpha ، ن - ١)} = كا^2_{(٠,٠٥ ، ٢٤)} = ٣٦,٤١٥$

(كما هو موضح بالشكل (٦ ــ ٤))



شكل (٦ ــ ٤)

وباستخدام بيانات العينة نحصل على :

$$كا^2 = \frac{ع^2 (ن - 1)}{\sigma_0^2} = \frac{(64)(24)}{80} = 19.2$$

وحيث أن قيمة كا^2 المحسوبة (١٩,٢) تقع في منطقة القبـــول فإننـــا نستنتج أن بيانات العينة غير كافية لتأييد الفرض البديل وذلـــك عنـــد مســـتوى معنوية ٠,٠٥ .

مثال (١١) :

يعتبر صافي حساب النقدية اليومية للشركة من الأمور التي تســـترعي اهتمام مجلس الإدارة . وكان رئيس مجلس إدارة إحدى الشركات يرى ضرورة دراسة تباين صافي حساب النقدية اليومية للشركة والذي أثبتت الدراسات السابقة أنه يساوي ١٠٠ ألف جنيه . في حين أنه كان يعتقد أن هـــذا التبـــاين يختلـــف بدرجة كبيرة عن هذه القيمة . وللتحقق من ذلك قام باختيـــار عينـــة عشـــوائية مكونة من صافي حساب النقدية اليومية لفترة حديثة شملت ١٤ يوما . وكـــانت النتائج كالموضحة بجدول (٦ ـــ ٢٣) هل تؤيد هذه البيانات اعتقـــاد رئيـــس مجلس الإدارة . استخدم مستوى معنوية ٠,١٠ .

جدول (٦ ــ ٢٣)

اليوم	صافي حساب النقدية اليومية (بآلاف الجنيهات)
١	١٧-
٢	٢٢+
٣	٥+
٤	صفر
٥	٨+
٦	٣-
٧	١٣+
٨	٢٠+
٩	٢٥+
١٠	٤١-
١١	صفر
١٢	٤-
١٣	٨+
١٤	١٣+

الحــــل :

حيث أننا نهدف لمعرفة ما إذا كان تباين صافي حساب النقدية اليوميــة للشركة يختلف عن ١٠٠ ألف جنيه فإن نوع الاختبار الذي يجب استخدامه هــو اختبار ذو طرفين وتكون عناصره كالآتي :

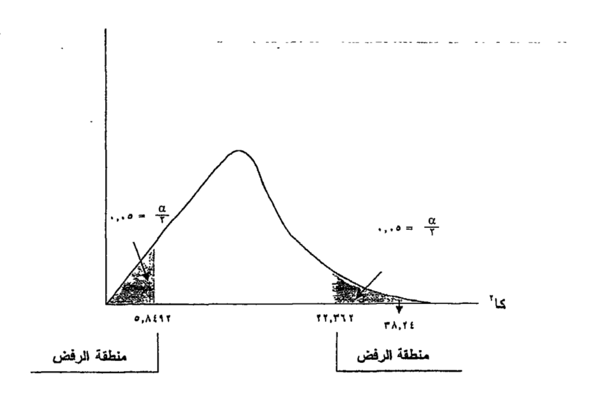
$H_0 : \sigma^2 = ١٠٠$ $\qquad$ $H_1 : \sigma^2 \neq ١٠٠$

إحصائية الاختبار :

$$كا^{٢} = \frac{ع^{٢} (ن - ١)}{\sigma_{٠}^{٢}}$$

منطقة الرفض كا^{٢} < كا$^{٢}_{(١ - \frac{\alpha}{٢} ، ن - ١)}$ أو كا^{٢} > كا$^{٢}_{(\frac{\alpha}{٢} ، ن - ١)}$

كا^{٢} < كا$^{٢}_{(٠,٩٥ ، ١٣)}$ = ٥,٨٩٢ أو كا^{٢} > كا$^{٢}_{(٠,٠٥ ، ١٣)}$ = ٢٢,٣٦٢



شكل (٦ ــ ٥)

ولحساب إحصائية الاختبار يجب أولا حساب تباين العينة ع^{٢} (كما هو موضح في جدول (٦ ــ ٢٤)) .

جدول (٦ ـــ ٢٤)

س	س - $\bar{س}$	$(س - \bar{س})^2$
١٧-	٢٠,٥-	٤٢٠,٢٥
٢٢+	١٨,٥	٣٤٢,٢٥
٥+	١,٥	٢,٢٥
صفر	٣,٥-	١٢,٢٥
٨+	٤,٥	٢٠,٢٥
٣-	٦,٥-	٤٢,٢٥
١٣+	٩,٥	٩٠,٢٥
٢٠+	١٦,٥	٢٧٢,٢٥
٢٥+	٢١,٥	٤٦٢,٢٥
٤١-	٤٤,٥-	١٩٨٠,٢٥
صفر	٣,٥-	١٢,٢٥
٤-	٧,٥-	٥٦,٢٥
٨+	٤,٥	٢٠,٢٥
١٣+	٩,٥	٩٠,٢٥
٤٩	صفر	٣٨٢٣,٥٠

$$\bar{س} = \frac{مجـ\ س}{ن} = \frac{٤٩}{١٤} = ٣,٥$$

$$ع^2 = \frac{مجـ\ (س - \bar{س})^2}{ن - ١} = \frac{٣٨٢٣,٥}{١٣} = ٢٩٤,١٢$$

$$\therefore\ كا^2 = \frac{ع^2\ (ن - ١)}{\sigma_0^2} = \frac{٢٩٤,١٢\ (١٣)}{١٠٠} = ٣٨,٢٤$$

وحيث أن كا^2 المحسوبة تقع في منطقة الرفض (كمـــا هـــو موضـــح بالشكل (٦ ـــ ٥)) فإننا نرفض فرض العدم . وبالتالي تؤيد البيانات اعتقـــاد رئيس مجلس الإدارة باختلاف تباين صافي حساب النقدية اليومية للشركة عـــن ١٠٠ ألف جنيه . واحتمال أن يكون هذا القرار خاطئاً ٠,١٠ .

تمارين (٦)

(١) أرادت إدارة إحدى الشركات التجارية معرفة الوسيلة الأكثر فاعلية في الإعلان عن الأوكازيون السنوي الذي تقوم به . فقامت باختيار عينة عشوائية من زبائن الأوكازيون وباستجوابهم عن الوسيلة التي أعلنوا بها عن الأوكازيون تم الحصول على النتائج الموضحة بالجدول التالي :

الوسيلة	التليفزيون	الراديو	الصحف والمجلات	الاتصال الشخصي بين الأفراد
عدد الزبائن	٥٣	٣٢	٣٦	٤٨

المطلوب :

أ ــ هل تؤيد هذه البيانات وجود اختلاف بين نسب الزبائن الذين أعلنوا بالوسائل الأربعة ؟ استخدم α = ٠,٠٥

ب ــ أوجد فترة ثقة ٩٠ ٪ لنسبة الزبائن الذين أعلنوا بالأوكازيون عن طريق الاتصال الشخصي بين الأفراد .

(٢) لمعرفة تأثير أحد الأدوية الجديدة في علاج مرض معين تم اختيار عينة عشوائية من ٢٠٠ مريض تناولوا هذا الدواء فوجد من بينهم ٩٠ مريض تحسنت حالتهم الصحية ، ٤٠ مريض لم تتغير حالتهم الصحية ، ٣٠ مريض حدثت لهم آثار جانبية ثانوية ، ٤٠ مريض حدثت لهم آثار جانبية رئيسية . فهل تتفق هذه البيانات مع توقع الشركة المنتجة للدواء بأن ٥٠ ٪ من المستخدمين للدواء سوف تتحسن حالتهم ، ٣٠ ٪ لن تتغير حالتهم الصحية ، ١٥ ٪ تحدث لهم آثار جانبية ثانوية ، ٥ ٪ تحدث لهم آثار جانبية رئيسية ، استخدم α = ٠,٠١ .

٣) أرادت منظمة العمل دراسة مشكلة العمال في ثلاث صناعات رئيسية حدث فيها إحلال الآلات محل العمل اليدوي . فقامت باختيار عينة عشوائية من مائة عامل من كل صناعة من الصناعات والذين فقدوا عملهم بسبب التقدم التكنولوجي وإحلال الآلات محل العمل اليدوي.وتم استجواب كل عامل عما إذا كان قد وجد عمل آخر داخل نفس الشركة أو في شركة جديدة وفي نفس الصناعة أو في صناعة جديدة أو لم يجد أي عمل آخر ، والجدول التالي يلخص النتائج التي تم الحصول عليها .

الوضع الحالي للعامل / الصناعة	نفس الشركة	شركة جديدة (نفس الصناعة)	صناعة جديدة	لم يجد عمل
أ	٦٢	١١	٢٠	٧
ب	٤٥	٨	٣٨	٩
جـ	٦٨	١٩	٨	٥

هل تؤيد هذه البيانات وجود علاقة بين الوضع الحالي للعامل والصناعة التي كان يعمل بها وتركها نتيجة للتقدم التكنولوجي وإحلال الآلات محل العمل اليدوي ؟ استخدم مستوى معنوية α = ٠,٠١ .

ـ التالي يبين توزيع مائة شخص أصيبوا بأزمة قلبية حسب الجنس

.

الجنس / العمر	ذكر	أنثى	المجموع
أقل من ٣٠ سنة	٦	٤	١٠
٣٠ – ٦٠	٣٨	٤٢	٨٠
أكثر من ٦٠	٦	٤	١٠
المجموع	٥٠	٥٠	١٠٠

هل تؤيد هذه البيانات وجود علاقة بين نوع الشخص المصاب بأزمة قلبية وعمره . استخدم مستوى معنوية α = ٠,١٠ .

(٥) قامت إحدى شركات الصناعات الدوائية بتحضير نوع جديد من الأدوية لعلاج الأرق ، وأرادت أن تقدر تباين الوقت الذي يمر بين تناول المريض للدواء ، وشعوره بالراحة والاسترخاء . وتم تجربة هذا الدواء على ٢٠ مريض .فوجد أن الانحراف المعياري للوقت الذي يمر بين تناول المريض للدواء وشعوره بالراحة والاسترخاء يساوي ١١ دقيقة.أوجد فترة ثقة ٩٠ ٪ لتباين الوقت الذي يمر بين تناول المريض للدواء وشعوره بالراحة والاسترخاء .

(٦) يستخدم أحد منتجي الغلال آلة معينة لتعبئة منتجه على أساس أن يكون الانحراف المعياري لوزن العبوة يساوي ١٠ جرام.ولكن لاحظ المنتج أن الانحراف المعياري لوزن بعض العبوات يختلف عن ١٠ جرام . فقام باختيار عينة عشوائية من ١٥ عبوة وحسب الانحراف المعياري فكان مساويا ١٥ جرام .

المطلوب :

أ ـــ إيجاد فترة ثقة ٩٥ ٪ لتباين وزن العبوة .

ب ـــ هل تؤيد هذه البيانات رأي المنتج بأن الانحراف المعياري لوزن العبوة يختلف عن ١٠ جرام . استخدم مستوى معنوية ١٠ ٪ .

(٧) تعاقد أحد مراكز البحوث على استيراد شحنة من الفئران وقد أعلن المصدر أن تباين وزن الفأر على الأكثر ٤ جرام فقام أحد الباحثين باختيار عينة عشوائية من ٧ فئران ووجد أن الانحراف المعياري لوزن

الفأر ٦ جرام . فهل تؤيد هذه البيانات إدعاء المصدر للفئران . استخدم مستوى معنوية α = ٠,٠٥ .

(٨) أراد أحد الطيارين دراسة توزيع عدد محركات الطائرة التي تحتاج إلى بنزين إضافي بعد ٧ ساعات طيران . فقام باختيار عينة عشوائية مكونة من ٢٠٠ رحلة طيران وسجل عدد المحركات التي تحتاج إلى بنزين إضافي بعد ٧ ساعات طيران في كل رحلة ، والجدول التالي يلخص النتائج التي تم الحصول عليها :

عدد المحركات التي تحتاج إلى بنزين بعد ٧ ساعات طيران	صفر	١	٢	٣	٤	المجموع
عدد رحلات الطيران	١٥	٤٤	٧٥	٥٦	١٠	٢٠٠

المطلوب : اختبار الفرض بأن عدد محركات الطائرة التي تحتاج إلى بنزين بعد ٧ ساعات طيران يتبع لتوزيع ذي الحدين باحتمال نجاح θ = ٠,٠٥ . استخدم مستوى معنوية α = ٠,٠١ .

(٩) أراد مراقب جودة الإنتاج أن يختبر فرض العدم بأن عدد الوحدات المعيبة في صندوق يحتوي على ٣ وحدات يتبع لتوزيع ذي الحدين باحتمال نجاح = ٠,٢ فقام باختيار عينة عشوائية من ٢٠٠ صندوق والجدول التالي يلخص النتائج التي تم الحصول عليها :

عدد الوحدات المعيبة في الصندوق	صفر	١	٢	٣	المجموع
عدد الصناديق	٤٦	٦٠	٥٨	٣٦	٢٠٠

المطلوب : إجراء الاختبار عند مستوى معنوية α = ٠,١٠ .

(١٠) أرادت هيئة النقل والمواصلات دراسة توزيع عـــدد الحـــوادث التـــي يرتكبها سائق الأتوبيس سنويا فقامت باختيار عينة عشوائية مكونة مـــن ١١٢٩٥٦ سائق وتم تسجيل عدد الحوادث التي ارتكبها السائق في الســنة وكانت كالمبينة بالجدول التالي :

عدد الحوادث التي ارتكبها السائق في السنة	صفر	١	٢	٣	المجموع
عدد السائقين	١٠٣٦٢٨	٧٣٨٩	١٩١٠	٢٩	١١٢٩٥٦

المطلوب :- اختبار فرض العدم بأن عدد الحوادث التي يرتكبها السائق في السنة يتبع لتوزيع بواسون . استخدم مستوى معنوية α = ٠,٠٥ .

(١١) الجدول التالي يبين توزيع عدد الأخطاء في الســـاعة لثلاثيـــن ســـاعة طيران :

عدد الأخطاء في الساعة	عدد الساعات
صفر	٣
١	٨
٢	٥
٣	٧
٤	٢
٥	١
٦	٢
٧	١
٨	صفر
٩	صفر
١٠	١
المجموع	٣٠

المطلوب : اختبار فرض العدم بأن هذه البيانات تتبع لتوزيع بواســـون . استخدم مستوى معنوية α = ٠,٠١ .

(١٢) في إحدى الدراسات التسويقية اختيرت عينة عشوائية مكونة مـــن ٥٠٠ مستهلك لأنواع مختلفة من الشاي وتم استجواب كل منهم عن النوع الأكثر تفضيلا لديه والجدول التالي يلخص نتائج الدراسة :

النوع الأكثر تفضيلا	عدد المستهلكين
أ	٩١
ب	١٠٩
جـ	٨٥
د	١٠٠
هـ	١١٥
المجموع	٥٠٠

اختبر فرض العدم بأن تفضيلات المستهلك للأنواع المختلفة من الشـــــاي تتبع للتوزيع المنتظم . استخدم مستوى معنوية α = ٠,١٠ .

(١٣) ألقيت زهرة نرد ١٢٠ مرة وكان عدد مرات ظهور الست أوجه مـــن ١ إلى ٦ على الترتيب ٥ ، ٣٦ ، ٨ ، ٩ ، ٧ ، ٥٥ . المطلـــــوب اختبـــار فرض العدم بأن الست أوجه في كل الرميات الممكنة لزهرة النرد تتـــوزع بانتظام . استخدم مستوى معنوية α = ٠,٠٥ .

(١٤) أراد أحد الباحثين الاقتصاديين دراسة التوزيع الاحتمالي للدخل الشهري لمهندسي المباني فقام باختيار عينة عشوائية مـــن ١٠٠٠ مـــهندس وتـــم تسجيل الدخل الشهري لكل منهم ، والجدول التالي يلخص النتائج التي تـــم الحصول عليها :

فئات الدخل الشهري	أقل من ٨٠٠	٨٠٠ –	٨٥٠ –	٩٠٠ –	٩٥٠ –	١٠٠٠ فأكثر	المجموع
عدد المهندسين	٢٦	١٤٦	٣٦١	٣١١	١٤٣	١٣	١٠٠٠

اختبر فرض العدم بأن الدخل الشهري لمهندسي المباني يتبع لتوزيع معتدل بوسط حسابي μ = ٩٠٠ وانحراف معياري σ = ٥٠ .

الفصل السابع

الارتباط الخطي بين الظواهر

٧-١ مقدمة:

تعتبر مقاييس الارتباط والتي نقدمها في هذا الفصل من الأدوات الهامة التي تجيب على العديد من الأسئلة المتعلقة بطبيعة العلاقات بين المتغيرات، وسوف تقتصر مناقشتنا في هذا الفصل على قياس قوة العلاقة الخطية بين متغيرين، وبالتالي سوف نقوم باستخدام تعبير "الارتباط الخطي بين المتغيرات".

٧-٢ معاملات الارتباط وخصائصها

يتم تحليل الارتباط الخطي دائماً على أساس حساب ما يسمى بمعامل الارتباط والذي نرمز له بالرمز "ر"، ويتصف معامل الارتباط بأن قيمته المطلقة لا تتجاوز الواحد الصحيح،

$$| ر | \leq ١$$

أو بمعنى آخر

$$-١ \leq ر \leq ١$$

ونستخلص من قيمة معامل الارتباط ما يلي:

1- إذا كانت قيمة معامل الارتباط تساوي صفر أو قريبة من الصفر فإننا نستنتج عدم وجود علاقة خطية بين المتغيرين، ويلاحظ هنا أننا ننفي

وجود العلاقة الخطية لأنه قد توجد حالات نجد فيها أن ر= صفر بينما توجد علاقة غير خطية بين المتغيرين كما سيتضح من الأمثلة فيما بعد.

2- إذا كانت إشارة معامل الارتباط موجبة فإن هذا يعني وجود علاقة خطية طردية بين المتغيرين، وإذا كانت إشارته سالبة دل ذلك على وجود علاقة عكسية بين المتغيرين.

3- إذا كانت ر = ١ أو -١ دل ذلك على وجود علاقة خطية تامة بين المتغيرين وهي الحالة التي نجد فيها أن جميع النقاط تقع على استقامة واحدة كما أسلفنا الذكر في الفصل السابق.

4- كلما اقتربت قيمة معامل الارتباط من الواحد الصحيح كلما زادت قوة العلاقة بين المتغيرين وكلما بعدت عن الواحد الصحيح واقتربت من الصفر ضعفت العلاقة بين المتغيرين، وبصورة تقريبية يمكن القول بأن العلاقة تعتبر قوية إذا زادت القيمة العددية لمعامل الارتباط عن ٠,٨ وتعتبر العلاقة متوسطة إذا انحصرت القيمة العددية لمعامل الارتباط بين ٠,٥ و ٠,٨ وتكون ضعيفة إذا قلت عن ٠,٥.

ونتناول فيما يلي بعض مقاييس معاملات الارتباط والتي تناسب الأنواع المختلفة للمتغيرات الإحصائية.

٧-٣ معامل بيرسون للارتباط

يستخدم معامل بيرسون لقياس الارتباط الخطي بين المتغيرات الكمية ولا يصلح للاستخدام في حالة البيانات النوعية، وتكون إحدى الصيغ الممكنة لحساب معامل بيرسون للارتباط الخطي بين متغيرين على الصورة

$$ر = ١ - \frac{١}{٢(ن-١)} \text{مجـ} (Z_س - Z_ص)^٢ \qquad (٧_١)$$

حيث تعبر $Z_س$ و $Z_ص$ عن الوحدات المعيارية المناظرة لمشاهدات كل من المتغيرين س وص على الترتيب وتعرف على الصورة:

$$Z_س = \frac{(س - \overline{س})}{ع_س} \qquad (٧_٢)$$

$$Z_ص = \frac{(ص - \overline{ص})}{ع_ص} \qquad (٧_٣)$$

وعند حساب الوحدات المعيارية لقيم أي متغير، سنجد أن وسطها الحسابي دائما ما يساوي الصفر وتباينها يساوي الواحد الصحيح، وبالتالي فهي تستخدم لمقارنة قيم الظواهر المختلفة على أساس "معياري" موحد يستبعد تأثير الفروق بين المستوى العام لقيم المتغيرات (بجعل متوسطاتها الحسابية تساوي الصفر)، ويستبعد كذلك تأثير اختلاف درجات تباين قيم كل متغير (بجعل انحرافاتها المعيارية تساوي الواحد الصحيح).

ويلاحظ القارئ هنا أننا استخدمنا أدلة سفلية لكل من الوحدات المعيارية والانحرافات المعيارية وذلك لبيان ما إذا كان المقياس المشار إليه يخص المتغير س أم المتغير ص.

ويمكن إعادة كتابة المعادلة (٧_١) على الصورة

$$ر = ١ - \frac{١}{٢(ن-١)} \left[\text{مجـ} Z_س^٢ + ٢ \text{ مجـ} Z_س Z_ص + \text{مجـ} Z_ص^٢ \right]$$

وحيث أن

$$\frac{\text{مجـ } Z_{س}^{2}}{ن - 1} = \frac{\text{مجـ }(س - \overline{س})^{2}}{ع_{س}^{2}(ن - 1)} = \frac{ع_{س}^{2}}{ع_{س}^{2}} = 1$$

وبنفس المنطق نجد أن

$$\frac{\text{مجـ } Z_{س}}{ن - 1}$$

وبالتالي تكون

$$ر = 1 - \frac{1}{2}\left[2 - 2\,\frac{\text{مجـ } Z_{س} Z_{ص}}{ن - 1}\right]$$

$$ر = \frac{\text{مجـ } Z_{س} Z_{ص}}{ن - 1} \qquad (٧_٤)$$

و بالتعويض عن $Z_{س}$ و $Z_{ص}$ من المعادلتين (٧_٢) و (٧_٣) في المعادلة (٧_٤) يمكن كتابة معامل بيرسون للارتباط على الصورة

$$ر = \frac{\frac{1}{(ن - 1)} مجـ (س - \overline{س})(ص - \overline{ص})}{ع_س ع_ص} \qquad (٧_٥)$$

ويطلق على البسط في الصيغة السابقة اسم التغاير والذي نرمز له بالرمز $ع_{س ص}$ ، بالتالي يمكن كتابة معامل بيرسون للارتباط على الصورة

$$ر = \frac{ع_{س ص}}{ع_س ع_ص} \qquad (٧_٦)$$

ويمكن اشتقاق صورة بديلة لمعامل بيرسون للارتباط وذلك بإعادة كتابة صيغة التغاير كما يلي

$$ر = \frac{مج (س - \overline{س})(ص - \overline{ص})}{ع_س ع_ص}$$

ويمكن إثات أن

$$ر = \frac{1}{ن - 1} \left(مج\, س ص - \frac{(مج\, س)(مج\, ص)}{ن} \right) \qquad (٧_٧)$$

ولتسهيل إجراء العمليات الحسابية في حالة ما إذا كانت قيم أحد المتغيرين أو كلاهما كبيرة، نناقش الآن تأثير إضافة مقدار ثابت وكذلك تأثير الضرب في مقدار ثابت على القيم المعيارية وبالتالي على قيمة معامل الارتباط.

افترض أننا قمنا بإضافة مقدار ثابت أ على كل قيمة من قيم س ثم ضربنا الناتج في مقدار ثابت ل بحيث نحصل على وحدات متغير جديد ولـيكن ي والذي يأخذ الصورة

ي = ل (س + أ)

وتكون مشاهدات المتغير الجديد ي على الصورة

$ي_1$ = ل ($س_1$ + أ)

$ي_2$ = ل ($س_2$ + أ)

$ي_3$ = ل ($س_3$ + أ)

. . . ،

. . . ،

$ي_ن$ = ل ($س_ن$ + أ)

والأن إذا أردنا حساب الوسط الحسابي لقيم المتغير ي فإننا نقـوم بأخـذ مجموع المعادلات السابقة، ونلاحظ هنا أنه عند أخذ مجموع الطرف الأيسر من المعادلات أن المقدار ل يكون عاملا مشتركا وبالتـالي نحصـل علـى الصيغة

مجـ ي = ل مجـ (س + أ)

مجـ ي = ل (مجـ س + مجـ أ)

وحيث أن أ مقدار ثابت وأن مجموع المقدار الثابت يساوي الثابـت مضروبا في عدد الحدود، فإننا نحص على الصيغة

مجـ ي = ل (مجـ س + ن أ)

بالتالي لحساب الوسط الحسابي لقيم ي نقسم الطرفين على ن لنحصل على العلاقة

$$\frac{مجـ\ ي}{ن} = ل\left(\frac{مجـ\ س}{ن} + \frac{ن\ أ}{ن}\right)$$

$$\overline{ي} = ل\,(\overline{س} + أ)$$

وتعني هذه النتيجة أنه للحصول على الوسط الحسابي لقيم ي نقوم بإجراء نفس التحويلات السابقة التي أجريناها على قيم س ونطبقها على الوسط الحسابي $\overline{س}$ ، أي نقوم بإضافة أ ثم الضرب في ل،

مثال ٧-١

إذا كان الوسط الحسابي لقيم المتغير س هو $\overline{س}=23$ ، أوجد الوسط الحسابي للمتغير ي والذي يعرف على الصورة

$$ي = ٠,٢(س + ٧)$$

الحل:

من العلاقة السابقة يكون الوسط الحسابي للمتغير ي على الصورة

$$\overline{ي} = ٠,٢\ (\overline{س} + ٧)$$

$$\overline{ي} = ٠,٢\ (٢٣ + ٧)$$

$$\overline{ي} = ٦\ .$$

يلاحظ في هذا المثال أن قيمة الوسط الحسابي للمتغير الجديد ي تتأثر بقيمتي الثابتين أ و ل.

وإذا أردنا دراسة العلاقة بين تباين قيم ي وتباين قيم س، فإننا نكتب أولا صيغة تباين قيم ي ثم نقوم بالتعويض بدلالة قيم س،

$$ع_ي^٢ = \frac{١}{ن-١} \text{ مجـ } (ي - \overline{ي})^٢$$

$$ع_ي^٢ = \frac{١}{ن-١} \text{ مجـ } (ل(س+أ) - ل(\overline{س}+أ))^٢$$

$$ع_ي^٢ = \frac{١}{ن-١} \text{ مجـ } (ل س - ل \overline{س})^٢$$

$$ع_ي^٢ = ل^٢ \frac{١}{ن-١} \text{ مجـ } (س - \overline{س})^٢$$

$$ع_ي^٢ = ل^٢ ع_س^٢$$

وبالتالي يكون الانحراف المعياري لقيم ي على الصورة

$$ع_ي = ل ع_س$$

أي أن الانحراف المعياري للمتغير الجديد ي لا يتأثر بعملية الجمع (لا يعتمد على الثابت الجمعي أ) ولكنه يتأثر بعملية الضرب (يعتمد على ل).

مثال ٧-٢

أوجد قيمة الانحراف المعياري للمتغير ي في المثال السابق إذا علمت أن الانحراف المعياري للمتغير س هو ١٠ .

الحل:

حيث أن

$$ي = ٠,٢(س + ٧)$$

فإن الانحراف المعياري لقيم ي يكون

$$ع_ي = ٠,٢ ع_س$$

$$= ٠,٢ \times ١٠$$

$$= ٢ .$$

لاحظ هنا أننا تجاهلنا ثابت الجمع، ٧، تماما.

والآن نقوم بدراسة تأثير عمليتي جمع وضرب مقدار ثابت، على قيم متغير، على الوحدات المعيارية له وذلك بإيجاد العلاقة بين $Z_س$ و $Z_ي$ ،

$$Z_ي = \frac{ي - \overline{ي}}{ع_ي}$$

$$= \frac{ل(س + أ) - ل(\overline{س} + أ)}{ل ع_س}$$

$$= \frac{ل س - ل\overline{س}}{ل ع_س}$$

$$= Z_س$$

وتعني النتيجة السابقة أن القيم المعيارية لمتغير معين لا تتأثر بعمليات جمع مقدار ثابت على قيم متغير معين أو ضربها في مقدار ثابت، وحيث أن صيغة معامل بيرسون للارتباط في المعادلة (٧_١) تعتمد فقط على القيم المعيارية للمتغيرين فإننا نستنتج أن قيمة معامل الارتباط هي الأخرى لن تتأثر بعمليات الجمع والضرب، بالتالي إذا وجدنا أن قيم أحد المتغيرين أو كلاهما كبيرة مما يجعل العمليات الحسابية معقدة فيمكن أن نقوم بطرح مقدار ثابت من قيم كل من المتغيرين، وإذا احتوت القيم على عامل مشترك فيمكن أيضا القسمة عليه ولن تختلف قيمة معامل الارتباط عند حسابها من القيم الأصلية أو القيم الجديدة.

لعل القارئ يتساءل عن سبب تعدد صيغ حساب معامل بيرسون للارتباط، يرجع السبب في ذلك إلى أن هذه الصيغ تخدم أغراضا مختلفة، فالصيغة المعطاة في المعادلة (٧_١) تستخدم في تفسير معنى هذا المعامل في ظل الدرجات المختلفة للعلاقات الخطية بين المتغيرين، أما الصيغة (٧_٦) فتكون مناسبة عند تحليل العلاقة بين الارتباط والانحدار الذي نقوم بدراسته في الفصل القادم، وحيث أن حساب الوحدات المعيارية لقيم المتغيرين ينطوي على عمليات حسابية مطولة، فإننا عادة ما نستخدم أحد الصيغتين (٧_٥) او (٧_٨) لحساب قيمة معامل بيرسون للارتباط، وننصح باستخدام الصيغة (٧_٥) إذا وجدنا أن قيمتي الوسط الحسابي للمتغيرين س وص عددين صحيحين ونقوم باستخدام الصيغة (٧_٨) فيما عدا ذلك.

نبدأ فيما يلي بعرض بعض الأمثلة لإرساء المفاهيم السابقة وبيان الخطوات المتبعة لحساب قيمة معامل بيرسون للارتباط، وسوف نقوم في الأمثلة الثلاثة الأولى بتطبيق الصيغة (٧_١) لإبراز بعض الحالات الهامة.

مثال ٧-٣

احسب معامل بيرسون للارتباط باستخدام المشاهدات التالية عن متغيرين س وص،

س	2	10	7	4	7	6
ص	7	23	17	11	17	15

الحل:

لحساب القيم المعيارية نبدأ أولا بحساب الوسط الحسابي والانحراف المعياري لكل من قيم س وص،

مسلسل	س	ص	س-٦	ص-١٥	$(س-٦)^2$	$(ص-١٥)^2$
١	٢	٧	٤-	٨-	١٦	٦٤
٢	١٠	٢٣	٤	٨	١٦	٦٤
٣	٧	١٧	١	٢	١	٤
٤	٤	١١	٢-	٤-	٤	١٦
٥	٧	١٧	١	٢	١	٤
٦	٦	١٥	٠	٠	٠	٠
المجموع	٣٦	٦٠	صفر	صفر	٣٦	١٥٦

$\overline{س} = ٣٦ \div ٦ = ٦$

$\overline{ص} = ٩٠ \div ٦ = ١٥$

$$ع_{س} = \sqrt{\frac{٣٨}{٥}} = ٢,٧٥٧$$

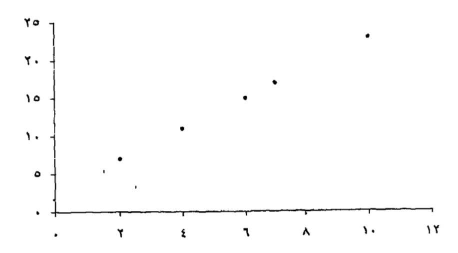
$$ع_{ص} = \sqrt{\frac{١٥٢}{٥}} = ٥,٥١٤$$

بقسمة قيم العمود الرابع في الجدول السابق على الانحراف المعياري لقيم س وقسمة قيم العمود الخامس على الانحراف المعياري لقيم ص نحصل على الوحدات المعيارية للمتغيرين كما في الجدول التالي

مسلسل	س	ص	س-٦	ص-١٥	$Z_{س}$	$Z_{ص}$
١	٢	٧	٤-	٨-	١,٦٤١-	١,٦٤١-
٢	١٠	٢٣	٤	٨	١,٦٤١	١,٦٤١
٣	٧	١٧	١	٢	٠,٣٦٢٧	٠,٣٦٢٧
٤	٤	١١	٢-	٤-	٠,٧٢٥٤-	٠,٧٢٥٤-
٥	٧	١٧	١	٢	٠,٣٦٢٧	٠,٣٦٢٧
٦	٦	١٥	٠	٠	٠	٠
المجموع	٣٦	٦٠	صفر	صفر	صفر	صفر

حيث أن جميع قيم الوحدات المعيارية للمتغيرين متساوية، فإن جميع مربعات الفروق في الصيغة (٧-١) سوف تكون مساوية للصفر وبالتالي تكون قيمة معامل بيرسمون للارتباط ، ر= ١.

ويرجع تساوي قيم الوحدات المعيارية للمتغيرين في هذا المثال إلى وجود علاقة خطية وطردية تامة بين المتغيرين مما يعني أن قيم أحد المتغيرين قد ضربت في مقدار ثابت وأضيف إليها ثابت آخر، وهذا كما سبق وأن رأينا لا يؤثر على القيم المعيارية، ويوضح شكل الانتشار التالي اتجاه العلاقة



نخلص من هذا إلى أنه إذا كانت .هناك علاقة خطية وطردية تامة بين المتغيرين، فإن قيم وحداتهما المعيارية سوف تكون دائما متساوية ويترتب على ذلك من المعادلة (٧_١) أن قيمة معامل الارتباط سوف تكون دائما مساوية للواحد الصحيح،

مثال ٧-٤

باستخدام المعادلة (٧_١) احسب معامل بيرسون للارتباط بين المتغيرين س وص من البيانات الآتية

س	1	5	5	1	8
ص	14	16	16	14	0

الحل:

نترك للطالب على سبيل التدريب أن يقوم بحساب قيمة كل من الوسط الحسابي والانحراف المعياري لقيم المتغيرين ليجد أن،

$\overline{س} = ٤$ $\overline{ص} = ٨$ $ع_{س} = ٣$ $ع_{ص} = ٦$،

ونقوم في الجدول التالي بحساب القيم المعيارية ومربعات الفروق بينها تمهيدا لاستخدام المعادلة (٧_١) لحساب قيمة ر.

س	ص	$Z_س$	$Z_ص$	$Z_س - Z_ص$	$(Z_س - Z_ص)^2$
1	١٤	١-	١	٢-	٤
٥	٦	٠,٣٣٣٣	٠,٣٣٣٣-	٠,٦٦٦٧	٠,٤٤٤٤٤
٥	٦	٠,٣٣٣٣	٠,٣٣٣٣-	٠,٦٦٦٧	٠,٤٤٤٤٤
١	١٤	١-	١	٢-	٤
٨	٠	١,٣٣٣٤	١,٣٣٣٤-	٢,٦٦٦٦	٧,١١١٢
٢٠	٤٠	صفر	صفر	صفر	١٦

من معادلة (٧ـ١) نجد أن

$$ر = ١ - \frac{١٦}{٢ \times ٤}$$

$$= ١ - ٢$$

$$= -١ ،$$

وتوضح هذه النتيجة أن هناك علاقة عكسية تامة بين المتغيرين، ويظهر من الجدول السابق أن هذه الحالة سوف تحدث عندما تتساوى القيم العددية للوحدات المعيارية وتختلف إشاراتها،

فإذا كانت $Z_ص = - Z_س$ لجميع القيم فإنه يمكن كتابة المعادلة (٧ـ١) على الصورة

$$ر = ١ - \frac{١}{٢(ن-١)} \text{ مجـ } (Z_س - (-Z_س))^٢$$

$$ر = ١ - \frac{٤}{(ن-١)^{٢}} مجـ (Z_{س})^{٢}$$

وحيث أن

$$\frac{مجـ Z_{س}^{٢}}{(ن-١)} = ١$$

فإن

$$ر = ١ - ٢ = -١$$

من المثالين السابقين يمكن أيضا استنتاج أنه كلما اقتربت القيم المعيارية للمتغيرين من بعضها البعض كلما قلت الفروق بينهـا واقتـرب مجمـوع المربعات في المعادلة (٧_١) من الصفر وبالتالي اقتربـت قيمـة معامـل الارتباط من الواحد الصحيح، ومن ناحية أخرى كلما اقتربت القيم المطلقة للوحدات المعيارية وكانت إشاراتها مختلفة كلما اقتربت قيمة معامل الارتباط من سالب واحد صحيح ليدل على وجود علاقة عكسية قوية بين المتغيرين،

مثال ٧-٥

استخدم المعادلة (٧_١) لحساب معامل الارتباط بين المتغيرين س وص من البيانات الآتية

س	3	5	4	8	10
ص	12	8	10	6	13

الحل:

عند حساب الوسط الحسابي والانحراف المعياري لكل من المتغيرين نجد أن

$\overline{س}$ = ٦ $\overline{ص}$ = ٩,٨ $ع_{س}$ = ٢,٩٢ $ع_{ص}$ = ٢,٨٦،

وبالتالي تكون الوحدات المعيارية ومربعات الفروق بينها كما في الجــدول التالي

س	ص	$Z_{س}$	$Z_{ص}$	$Z_{س}-Z_{ص}$	$(Z_{س}-Z_{ص})^{٢}$
٣	١٥	١,٠٢٨٩٩-	٠,٧٦٨٢٧	١,٧٩٧٢٦-	٣,٢٣٠١٦
٥	٨	٠,٣٤٣٠٠-	٠,٦٢٨٥٩-	٠,٢٨٥٥٩	٠,٠٨١٥٦
٤	١٠	٠,٦٨٥٩٩-	٠,٠٦٩٨٤	٠,٧٥٥٨٤	٠,٥٧١٢٩
٨	٦	٠,٦٨٥٩٩	١,٣٢٧٠٢-	٢,٠١٣٠١	٤,٠٥٢٢٢
١٠	١٣	١,٣٧١٩٩	١,١١٧٤٩	٠,٢٥٤٥٠	٠,٠٦٤٧٧
		صفر	صفر	صفر	٨,٠٠٠

من معادلة (٧_١) نجد أن

$$ر = ١ - \frac{١ \times ٨}{(٤)^{٢}} = ١-١ = صفر$$

أي أنه لا توجد علاقة خطية بين المتغيرين،

وعند النظر إلى قيم الوحدات المعيارية للمتغيرين نجد في هذه الحالة أن الإشارات تكون عكسية في بعض الأحيان وتكون متماثلة في أحيان أخرى، كذلك نجد أن بعض القيم المعيارية الكبيرة لأحد المتغيرين يناظرها قيم معيارية، أحيانا صغيرة وأحينا أخرى كبيرة للمتغير الأخر، ويعني هذا أن معامل الارتباط سوف يكون مساويا للصفر أو قريبا منه إذا كان هناك نمط عشوائي لارتباط القيم المعيارية ببعضها البعض وأيضا إشاراتها.

نتناول فيما يلي بعض الأمثلة لتوضيح كيفية حساب قيمة معامل بيرسون للارتباط باستخدام المعادلتين (٧_٥) و(٧_٨)

مثال ٧-٦

جمعت البيانات التالية عن درجات عينة من عشرة طلاب في امتحانين للمحاسبة والإدارة

رقم الطالب	محاسبة	إدارة		رقم الطالب	محاسبة	إدارة
1	72	78		6	54	52
2	76	82		7	58	66
3	64	78		8	72	74
4	68	60		9	90	95
5	81	70		10	75	75

والمطلوب:

1- رسم شكل الانتشار بين المتغيرين والتعليق عليه من حيث درجة الارتباط الخطي واتجاه العلاقة،

2- حساب قيمة معامل بيرسون للارتباط بين درجات الطلبة في امتحان المحاسبة وامتحان الإدارة،

الحل:

1- يوضح شكل الانتشار وجود لتجاه علاقة موجبة بين درجات الطلبة في المادتين ولكننا لا نتوقع هنا أن تكون قوية جدا، ويتضح ذلك من تباعد النقاط إلى حد ما، والأساس المنطقي الذي يمكن الاستناد إليه هو أن الطالب المتفوق يكون متفوقا في جميع الحالات والطالب محدود المستوى يكون أداءه على مستوى ضعيف في جميع المواد، ولكن من جهة أخرى قد يؤدي ميل بعض الطلبة نحو العلوم الإدارية إلى التفوق فيها والحصول على درجات مرتفعة في امتحاناتها أكثر مما يحصلون عليه في المواد الأخرى، ويحدث العكس في حالة ميل الطالب نحو العلوم المحاسبية، وتتسبب مثل هذه الميول المختلفة في انحراف القيم المشاهدة للدرجات عن الأساس المنطقي وبالتالي تضعف من قوة العلاقة بين الدرجات في المادتين .

لحساب معامل الارتباط باستخدام المعادلة (٧_٨) نعتبر أن س تمثل درجات الطلبة في المحاسبة وص تمثل درجاتهم في الإدارة، بعد ذلك نقوم بضرب كل قيمة من قيم س في قيمة ص المناظرة لها، وكذلك نقوم بإيجاد مربع كل قيمة من قيم س ومربع كل قيمة من قيم ص ثم نأخذ مجاميع الأعمدة المختلفة ونعوض بها في صيغة المعادلة،

رقم الطالب	س	ص	س ص	س٢	ص٢
١	٧٢	٧٨	٥٦١٦	٥١٨٤	٦٠٨٤
٢	٧٦	٨٢	٦٢٣٢	٥٧٧٦	٦٧٢٤
٣	٦٤	٧٨	٤٩٩٢	٤٠٩٦	٦٠٨٤
٤	٦٨	٦٠	٤٠٨٠	٤٦٢٤	٣٦٠٠
٥	٨١	٧٠	٥٦٧٠	٦٥٦١	٤٩٠٠
٦	٥٤	٥٢	٢٨٠٨	٢٩١٦	٢٧٠٤
٧	٥٨	٦٦	٣٨٢٨	٣٣٦٤	٤٣٥٦
٨	٧٢	٧٤	٥٣٢٨	٥١٨٤	٥٤٧٦
٩	٩٠	٩٥	٨٥٥٠	٨١٠٠	٩٠٢٥
١٠	٧٥	٧٥	٥٦٢٥	٥٦٢٥	٥٦٢٥
مجـ	٧١٠	٧٣٠	٥٢٧٢٩	٥١٤٣٠	٥٤٥٧٨

$$\text{ر} = \frac{\frac{1}{\text{ن}-1}\left[\text{مجـ س ص} - \frac{(\text{مجـ س})(\text{مجـ ص})}{\text{ن}}\right]}{\sqrt{\frac{1}{\text{ن}-1}\left[\text{مجـ س}^{٢} - \frac{(\text{مجـ س})^{٢}}{\text{ن}}\right]}\;\sqrt{\frac{1}{\text{ن}-1}\left[\text{مجـ ص}^{٢} - \frac{(\text{مجـ ص})^{٢}}{\text{ن}}\right]}}$$

$$ر = \frac{\frac{١}{٩}\left[٥٢٧٢٩ - \frac{(٧٣٠)\,(٧١٠)}{١٠}\right]}{\sqrt{\frac{١}{٩}\left[٥١٤٣٠ - \frac{(٧١٠)^{٢}}{١٠}\right]}\;\sqrt{\frac{١}{٩}\left[٥٤٥٧٨ - \frac{(٧٣٠)^{٢}}{١٠}\right]}}$$

$$ر = \frac{٩٩,٨٨٩}{١٠,٦٤٥ \times ١١,٩٦٣}$$

$$= ٠,٧٨٤٠$$

وتؤكد قيمة معامل بيرسون للارتباط النتيجة التي توقعناها من قبل وهي وجود علاقة طردية بين درجات الطلبة في المادتين ولكنها ليست عالية القوة.

يوضح المثال السابق حجم العمليات الحسابية التي نجريها أثناء إيجاد حواصل الضرب ومربعات الأرقام، وحتى إذا كنا نستخدم الآلات الحاسبة فإن تسجيل العديد من الأرقام يزيد من احتمالات الخطأ، كذلك يلاحظ أن الوسط الحسابي لكل من المتغيرين يكون عددا صحيحا وبالتالي فإن استخدام المعادلة (٧_٥) في إيجاد قيمة معامل الارتباط تعتبر أفضل كثيرا في هذه الحالة كما يتضح مما يلي،

إعادة حل المثال باستخدام المعادلة (٧_٥)

في هذه الحالة تكون الخطوات المتبعة في الحل كما يلي:

1- نقوم بحساب الوسط الحسابي للمتغيرين،

2- نوجد انحرافات قيم س عن س̄،

3- نوجد انحرافات قيم ص عن ص̄،

4- نوجد حواصل ضرب الانحرافات المتناظرة،

5- نقوم بإيجاد مربعات انحرافات قيم س وقيم ص،

6- نأخذ المجاميع المختلفة ونعوض في العلاقة

س̄ = ٧١ ص̄ = ٧٣

س	ص	س – س̄	ص – ص̄	(س – س̄)×(ص – ص̄)	(س – س̄)٢	(ص – ص̄)٢
٧٢	٧٨	١	٥	٥	١	٢٥
٧٦	٨٢	٥	٩	٤٥	٢٥	٨١
٦٤	٧٨	٧-	٥	٣٥-	٤٩	٢٥
٦٨	٦٠	٣-	١٣-	٣٩	٩	١٦٩
٨١	٧٠	١٠	٣-	٣٠-	١٠٠	٩
٥٤	٥٢	١٧-	٢١-	٣٥٧	٢٨٩	٤٤١
٥٨	٦٦	١٣-	٧-	٩١	١٦٩	٤٩
٧٢	٧٤	١	١	١	١	١
٩٠	٩٥	١٩	٢٢	٤١٨	٣٦١	٤٨٤
٧٥	٧٥	٤	٢	٨	١٦	٤
مجـ		٠	٠	٨٩٩	١٠٢٠	١٢٨٨

$$ر = \frac{\frac{1}{ن-1} \text{ مجـ}(س-\overline{س})(ص-\overline{ص})}{\sqrt{\frac{\text{مجـ}(س-\overline{س})^{2}}{ن-1}}\sqrt{\frac{\text{مجـ}(ص-\overline{ص})^{2}}{ن-1}}}$$

$$ر = \frac{٨٩٩}{\sqrt{١٠٢٠}\sqrt{١٢٨٨}}$$

$$ر = ٠,٦٨٤$$

يلاحظ أن كلا من البسط والمقام في صيغة الارتباط دائما ما يحتويان على نفس العامل (ن-١) وبالتالي يمكن إهماله عند الحساب كما فعلنا في الخطوة السابقة،

مثال ٧-٧

في دراسة للعلاقة بين طول لاعب كرة السلة وعدد الرميات الثلاثية التي يحرزها اللاعب جمعت البيانات التالية من عينة تتكون من ثمانية لاعبين حيث سجلت أطوالهم (س) وعدد الرميات الثلاثية التي أحرزها كل منهم في آخر أربع مباريات (ص)،

والمطلوب حساب معامل بيرسون للارتباط بين المتغيرين،

س	178	184	180	189	176	192	187	180
ص	5	7	6	9	4	10	8	9

الحل:

نقوم أولا بحل المثال باستخدام القيم الأصلية للمشاهدات، وبملاحظة أن قيم المتغير س كبيرة وأن وسطه الحسابي هو رقم كسري (١٨٣,٢٥) ، فإننا سوف نقوم بإعادة الحل بعد طرح مقدار ثابت من جميع قيم س وليكن (١٨٠) ، وحيث أن قيم ص صغيرة ولا توجد صعوبة في التعامل معها حسابيا فإننا نتركها كما هي،

س	ص	س ص	$س^2$	$ص^2$
١٧٨	٥	٨٩٠	٣١٦٨٤	٢٥
١٨٤	٧	١٢٨٨	٣٣٨٥٦	٤٩
١٨٠	٦	١٠٨٠	٣٢٤٠٠	٣٦
١٨٩	٩	١٧٠١	٣٥٧٢١	٨١
١٧٦	٤	٧٠٤	٣٠٩٧٦	١٦
١٩٢	١٠	١٩٢٠	٣٦٨٦٤	١٠٠
١٨٧	٨	١٤٩٦	٣٤٩٦٩	٦٤
١٨٠	٩	١٦٢٠	٣٢٤٠٠	٨١
١٤٦٦	٥٨	١٠٦٩٩	٢٦٨٨٧٠	٤٥٢

$$ع_{سص} = \frac{١}{٧}\left[١٠٦٩٩ - \frac{٥٨ \times ١٤٦٦}{٨}\right]$$

$$= ١٠,٠٧١٤$$

$$ع_{ص} = \sqrt{\frac{١}{٧}\left(٢٦٨٨٧٠ - \frac{(١٤٦٦)^{٢}}{٨}\right)}$$

$$= ٥,٦٧٦.$$

$$ع_{ص} = \sqrt{\frac{١}{٧}\left(٤٥٢ - \frac{(٥٨)^{٢}}{٨}\right)}$$

$$= ٢,١٢١$$

وحيث أن

$$ر = \frac{ع_{س ص}}{ع_{س}\, ع_{ص}}$$

فإن

$$ر = \frac{١٠,٠٧١٤}{٥,٦٧٦ \times ٢,١٢١}$$

$$ر = ٠,٨٣٧$$

يلاحظ في هذا المثال أننا قد قمنا بحساب مكونات صيغة معامل الارتباط كل على حدة ثم استخدمنا المعادلة (٧_٦) لحساب قيمته،

نقوم الآن بحل نفس المثال بطرح القيمة ١٨٠ كوسط فرضي من قيم س

$$ح_{س} = س - ١٨٠$$

ثم نجري نفس العمليات السابقة باستخدام عمودي $ح_{س}$ و ص لحساب قيم معامل الارتباط،

س	حس	ص	س ص	حس٢	ص٢
١٧٨	٢-	٥	١٠-	٤	٢٥
١٨٤	٤	٧	٢٨	١٦	٤٩
١٨٠	٠	٦	٠	٠	٣٦
١٨٩	٩	٩	٨١	٨١	٨١
١٧٦	٤-	٤	١٦-	١٦	١٦
١٩٢	١٢	١٠	١٢٠	١٤٤	١٠٠
١٨٧	٧	٨	٥٦	٤٩	٦٤
١٨٠	٠	٩	٠	٠	٨١
مجـ	٢٦	٥٨	٢٥٩	٣١٠	٤٥٢

$$\frac{\frac{١}{٧}\left[٢٥٩ - \frac{(٢٦)(٥٨)}{٨}\right]}{\sqrt{\frac{١}{٧}\left[٣١٠ - \frac{(٢٦)^{٢}}{٨}\right]}\ \sqrt{\frac{١}{٧}\left[٤٥٢ - \frac{(٥٨)^{٢}}{٨}\right]}}$$

$$ر = \frac{١٠,٠٧١}{٢,١٢١ \times ٥,٦٧٦}$$

$$ر = ,٨٣٧$$

من هذا يرى القارئ، وكما ذكرنا من قبل، أن قيمة معامل بيرسون للارتباط لا تتأثر عند طرح أو جمع مقدار ثابت على قيم أحد المتغيرين أو كلاهما، وتوضح قيمة معامل الارتباط وجود علاقة طردية قوية بين طول اللاعب وعدد الرميات الثلاثية التي ينجح في تصويبها،

سبق أن ذكرنا أن معامل بيرسون للارتباط يقيس قوة العلاقة الخطية بين متغيرين وبالتالي يمكن أن نجد أن قيمة معامل بيرسون للارتباط تساوي صفر أو تكون قريبة منه ومع ذلك تكون هناك علاقة شبه تامة بين المتغيرين ولكنها غير خطية كما يتضح من المثال التالي.

مثال ٧-٨

احسب معامل بيرسون للارتباط بين المتغيرين س و ص باستخدام البيانات التالية:

س	1	2	3	5	6	7
ص	11	6	3	3	6	11

الحل:

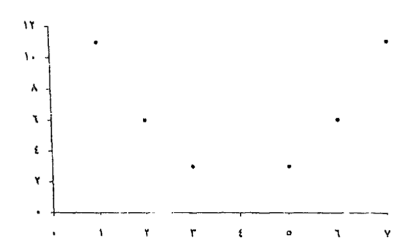
نقوم بحساب القيم المختلفة في الجدول التالي

س	ص	س ص	س٢	ص٢
١	١١	١١	١	١٢١
٢	٦	١٢	٤	٣٦
٣	٣	٩	٩	٩
٥	٣	١٥	٢٥	٩
٦	٦	٣٦	٣٦	٣٦
٧	١١	٧٧	٤٩	١٢١
٢٤	٤٠	١٦٠	١٢٤	٣٣٢

$$ر = \frac{١٦٠ - \frac{(٢٤)(٤٠)}{٦}}{\sqrt{١٢٤ - \frac{(٢٤)^{٢}}{٦}}\ \sqrt{٣٣٢ - \frac{(٤٠)^{٢}}{٦}}}$$

ر= صفر.

من ناحية أخرى، إذا نظرنا إلى شكل الانتشار نجده على صورة منحنى معادلة من الدرجة الثانية كما في الشكل التالي



في المثال التالي نوضح تأثير القيم المتطرفة على قيمة معامـل بيرسـون للارتباط وذلك بحساب قيمة المعامل من البيانات الكاملة ثـ حساب قيمتـه بعد حذفها،

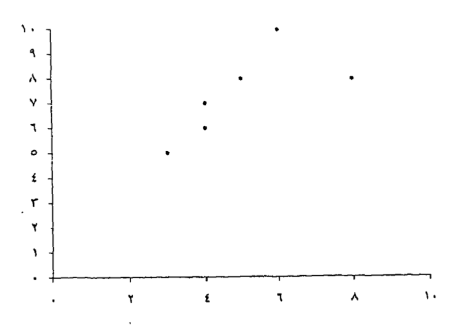
مثال ٧-٩

ارسم شكل الانتشار ثم احسب قيمة معامل بيرسون للارتبـــاط باســـتخدام البيانات التالية

س	3	4	4	4	5	6	8
ص	5	6	7	6	8	10	8

الحل :

نلاحظ من شكل الانتشار التالي أن جميع النقاط تقع إلى حد كبير في اتجاه خط مستقيم ما عدا المشاهدة الأخيرة حيث يكون مستواها منخفض بصورة واضحة عن اتجاه هذا الخط، بالتالي تعتبر المشاهدة الأخيرة قيمة شاذة عن الاتجاه الذي تأخذه باقي القيم،



س	ص	س ص	$س^{٢}$	$ص^{٢}$
٣	٥	١٥	٩	٢٥
٤	٦	٢٤	١٦	٣٦
٤	٧	٢٨	١٦	٤٩
٤	٦	٢٤	١٦	٣٦
٥	٨	٤٠	٢٥	٦٤
٦	١٠	٦٠	٣٦	١٠٠
٨	٨	٦٤	٦٤	٦٤
٣٤	٥٠	٢٥٥	١٨٢	٣٧٤

تكون قيمة معامل الارتباط في وجود القيمة المتطرفة

$$ر = \frac{٢٥٥ - \frac{(٣٤ \times ٥٠)}{٧}}{\sqrt{١٨٢ - \frac{(٣٤)^{٢}}{٧}} \sqrt{٣٧٤ - \frac{(٥٠)^{٢}}{٧}}}$$

$ر = ٠,٧٢٠٣$

وإذا قمنا بحذف المشاهدة الأخيرة ثم أعدنا تكوين جدول البيانات

س	ص	س ص	$س^{٢}$	$ص^{٢}$
٣	٥	١٥	٩	٢٥
٤	٦	٢٤	١٦	٣٦
٤	٧	٢٨	١٦	٤٩
٤	٦	٢٤	١٦	٣٦
٥	٨	٤٠	٢٥	٦٤
٦	١٠	٦٠	٣٦	١٠٠
٢٦	٤٢	١٩١	١١٨	٣١٠

سوف نجد أن قيمة معامل الارتباط

ر= ٠,٩٧٤

يوضح هذا المثال التأثير الشديد لوجود مشاهدة شاذة في البيانات على القيمة المحسوبة لمعامل الارتباط حيث زادت قيمته بعد استبعاد هذه المشاهدة من ٠,٧٢ إلى ٠,٩٧٤

٧-٤ معامل التحديد Coefficient of Determination

ذكرنا من قبل أن أحد أهداف تحليل العلاقات بين المتغيرات هو التعرف على دور المتغير أو المتغيرات المستقلة في تحديد قيم المتغير التابع وتفسير الاختلافات المشاهدة في قيمه، وعند دراسة العلاقة بين متغيرين فقط لا غير، فإننا نطلق على مربع قيمة معامل بيرسون للارتباط اسم معامل التحديد (ر^{٢})، ويقيس معامل التحديد نسبة الاختلافات المشاهدة في قيم المتغير التابع والتي يمكن تفسيرها من خلال تأثير المتغير المستقل عليه، فعلى سبيل المثال، إذا كانت قيمة معامل الارتباط بين المتغير التابع ص، والمتغير المستقل س، هي ٠,٩ ، فإن قيمة معامل التحديد تكون ٠,٨١ ، وتعني هذه القيمة أن ٨١% من الاختلافات المشاهدة في قيم ص تفسر باختلافات القيم المناظرة للمتغير المستقل س، وتمثل النسبة المتممة، ١٩%، تأثير بعض المتغيرات الأخرى التي تربطها علاقة بالمتغير التابع ص والتي لم تدخل في نموذج الدراسة، وفي حالة تحليل العلاقة بين متغير تابع وأكثر من متغير مستقل، فإننا نستخدم تعبير معامل التحديد المتعدد وسوف نشرح فيما بعد الأسلوب العام لحساب قيم معاملات التحديد بصرف النظر عن عدد المتغيرات المستقلة في النموذج، ونرمز لمعامل التحديد بين

المتغير التابع ص ومتغيرين مستقلين، س وع على سبيل المثال، بوضع رموز المتغيرات المستقلة كأدلة سفلية بين قوسين $ر^{٢}_{ص(س ع)}$، ويطلق على الجذر التربيعي لمعامل التحديد المتعدد اسم معامل الارتباط المتعدد والذي يقيس قوة العلاقة الخطية بين المتغير التابع ومجموعة المتغيرات المستقلة في النموذج، وفي الحالة الخاصة بدراسة العلاقة الخطية بين متغير تابع ص ومتغيرين مستقلين س و ع، يمكن حساب قيمة معامل التحديد المتعدد باستخدام معاملات بيرسون للارتباط بين كل متغيرين باستخدام العلاقة

معامل التحديد المتعدد =

$$\frac{ر^{٢}_{ص س} + ر^{٢}_{ص ع} - ٢ ر_{ص س} ر_{ص ع} ر_{س ع}}{١ - ر^{٢}_{س ع}}$$

معامل الارتباط المتعدد = $\sqrt{\text{معامل التحديد المتعدد}}$

مثال ٧-١٠

في نموذج لتحليل العلاقة الخطية بين متغير تابع ص ومتغيرين مستقلين س وع كانت معاملات الارتباط بين كل زوج من المتغيرات كما يلي:

$ر_{ص س} = ٠,٧٤٧$ $ر_{ص ع} = ٠,٦٤٨$ $ر_{س ع} = ٠,٧٦٤$

احسب قيمة كل من معامل التحديد المتعدد ومعامل الارتباط المتعدد،

الحل:

باستخدام الصيغة السابقة تكون قيمة معامل التحديد المتعدد

$r_{\text{ص س}} = ٠,٧٤٧ \qquad r_{\text{ص ع}} = ٠,٦٤٨ \qquad r_{\text{س ع}} = ٠,٧٦٤$

$$r^2_{\text{ص(س ع)}} = \frac{(٠,٧٤٧)^٢ + (٠,٦٤٨)^٢ - ٢ \times ٠,٧٤٧ \times ٠,٦٤٨ \times ٠,٧٦٤}{١ - (٠,٧٦٤)^٢}$$

$$= ٠,٥٧٢$$

$$r_{\text{ص(س ع)}} = \sqrt{٠,٥٧٢} = ٠,٧٥٧$$

تعني قيمة معامل الارتباط الكلي أنه يوجد اتجاه علاقة خطية بين المتغيرين ولكنها متوسطة وليست قوية، وإذا نظرنا إلى معامل التحديد المتعدد نجد أن المتغيرين المستقلين س و ع يفسران فقط ٥٧,٢% من تغيرات المتغير التابع ص، وإذا ما حسبنا نسبة تفسير كل متغير مستقل لتغيرات المتغير التابع على حده نجد أن $r^٢_{\text{ص(س)}}$= ٥٥,٨% وأن $r^٢_{\text{ص(ع)}}$=٤٢%، نستنتج من هذا أن نسبة تفسير المتغيرين المستقلين لتغيرات المتغير التابع ليست بصفة عامة محصلة مجموع نسبة تفسير كل متغير مستقل على حده، ولن يحدث هذا إلا إذا كانت قيمة معامل الارتباط بين المتغيرين المستقلين مساوية للصفر كما يتضح من الصيغة السابقة، ويمكن تفسير العبارة السابقة بأنه إذا لم تكن قيمة معامل الارتباط بين المتغيرين المستقلين مساوية للصفر، فإن هذا يتضمن وجود اتجاه مشترك لهما للتأثير على

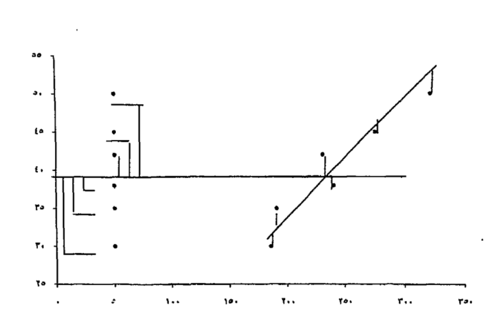
المتغير التابع بنفس النمط، ويظهر هذا الاتجاه المشترك في قيمة معامل التحديد الخاصة بكل متغير مستقل مع المتغير التابع، وبالتالي نجد أن مجموع قيمتي معامل التحديد تكون أكبر من قيمة معامل التحديد الكلي،
نقدم فيما يلي الأسلوب العام لحساب قيمة معامل التحديد المتعدد ونشرح كيفية النظر إلى قيمته كنسبة تفسير المتغيرات المستقلة لتغيرات قيم المتغير التابع وذلك من خلال تحليل المثال التالي.

مثال ٧-١١

القائمة التالية لأسعار بيع ستة كتب مختلفة،

من هذه البيانات نجد أن متوسط سعر الكتاب يكون، $\overline{ص}$ = ٤٠ ، وتكون انحرافات أسعار الكتب عن وسطها الحسابي على الصورة

السعر(ص)	30	35	38	42	45	50
ص- 40	10-	5-	2-	2	5	10



تظهر في الصف الثالث من الجدول السابق بمثابة أخطاء عشوائية، ولما كان مجموع هذه الانحرافات يساوي الصفر، فإننا نستخدم مجموع مربعاتها كمقياس لحجم الاختلافات الكلية (total variation)، وسوف نطلق على مجموع مربعات انحرافات القيم عن وسطها الحسابي اسم مجمع المربعات الكلي (sum of squares of total variation)، أي أن

مجموع المربعات الكلي = مجـ(ص− $\overline{ص}$)^{٢}

يلاحظ أن مجموع المربعات الكلي هو البسط في صيغة تباين ص،

وتظهر أحجام الاختلافات الكلية في شكل الانتشار السابق من خلال الأبعاد الرأسية للنقاط الموقعة إلى يمين المحور الرأسي مباشرة عن الخط الأفقي الذي يمثل قيمة الوسط الحسابي (٤٠)،

والآن إذا نظرنا إلى عدد صفحات الكتاب كأحد العوامل المؤثرة في سعر بيعه وكانت البيانات كما يلي

م	1	2	3	4	5	6
عدد الصفحات(س)	185	190	240	230	275	320
السعر(ص)	30	35	38	42	45	50

بالعودة إلى شكل الانتشار سنجد أن أدنى نقطة في النقاط الرأسية سوف تنتقل يمينا لتوقع أعلى القيمة ١٨٥، كذلك سوف تنتقل القيمة التالية (٣٥) لتوقع أعلى القيمة ١٩٠ وهكذا، مع أخذ عدد صفحات الكتاب في الاعتبار،

يبين شكل الانتشار اتجاه علاقة خطية طردية نستطيع من خلالها القول بأن سعر الكتاب الأول (٣٠) جاء منخفضا لأن عدد صفحاته هي الأقل، بينما سعر الكتاب السادس هو الأعلى لأن عدد صفحاته هي الأكبر، ولكون العلاقة غير تامة تظهر في الشكل انحرافات رأسية بين النقاط المشاهدة والنقاط المقابلة لها على خط العلاقة وهذه الانحرافات مرة أخرى تمثل أخطاء عشوائية لا يمكن تفسيرها في ظل غياب معلومات عن أية متغيرات مستقلة أخرى لها تأثير على سعر بيع الكتاب ونطلق عليها اسم التغيرات غير المفسرة (**unexplained variation**)، وإذا رمزنا للقيم المقدرة للسعر على خط العلاقة بالرمز صَ، فإن التغيرات غير المفسرة والتي تستخدم كتقدير للأخطاء العشوائية تحسب بالفروق

التغيرات غير المفسرة= خ̂ = (ص – صَ)

ويلاحظ أن أحجام الأخطاء في ظل النظر إلى أسعار الكتب من خلال أعداد صفحاتها أصبحت أقل كثيرا مما كانت عليه عند النظر إلى الأسعار فقط، وتمثل الفروق بين التغيرات الكلية والتغيرات غير المفسرة بعد أخذ عدد صفحات الكتاب في الاعتبار، مساهمة قيم المتغير المستقل في تفسير اختلافات قيم المتغير التابع ونطلق عليها اسم التغيرات المفسرة (**explained variation**)، أي أن

التغيرات المفسرة = التغيرات الكلية – التغيرات غير المفسرة

= (ص – $\overline{\text{ص}}$) – (ص – صَ)

= (صَ – $\overline{\text{ص}}$)

وإذا ما قمنا بحساب مجموع التغيرات المفسرة ومجموع التغيـرات غيـر المفسرة فسوف نجد أن كل منهما يساوي الصفر، بالتـالي نقـوم بأخـذ مجموع مربعات نوعي التغيرات السابقين لنحصل على

مجموع المربعات المفسر = مجـ $(\hat{ص} - \bar{ص})^2$

مجموع المربعات غير المفسر = مجـ $(ص - \hat{ص})^2$

وسوف نشاهد دائما صحة العلاقة التالية:

مجموع المربعات الكلي = مجموع المربعات المفسر + مجموع المربعات غير المفسرة

$$مجـ (ص - \bar{ص})^2 = مجـ (\hat{ص} - \bar{ص})^2 + مجـ (ص - \hat{ص})^2$$

والنسبة بين مجموع المربعات المفسر إلى مجموع المربعات الكلي تمثـل نسبة مساهمة المتغير المستقل في تفسير تغيـرات قـيم المتغيـر التـابع وبالتالي فإنها تعطي قيمة معامل التحديد، أي أنه يمكن حساب قيمة معامل التحديد باستخدام الصيغة

$$معامل\ التحديد = ر^2 = \frac{مجموع\ المربعات\ المفسر}{مجموع\ المربعات\ الكلي}$$

$$معامل\ التحديد = ر^2 = \frac{مجـ (\hat{ص} - \bar{ص})^2}{مجـ (ص - \bar{ص})^2}$$

يتضح من الصيغ السابقة ضرورة معرفة معادلة الخط المستقيم الذي يمثل العلاقة بين المتغير التابع والمتغيرات المستقلة وذلك حتـى نـتمكن مـن

التعويض في هذه المعادلة بقيم المتغير أو المتغيرات المستقلة لحساب القيم المقدرة، وسوف نتناول في الفصل القادم كيفية تقدير مثل هذه المعادلات،

دعنا نفترض الآن أن العلاقة المقدرة بين أسعار الكتب (ص) وعدد صفحاتها (س) تكون على الصورة:

صَ = ٨,٢١١٣ + ٠,١٣٥٥س

وبحساب معامل بيرسمون للارتباط بين المتغيرين نجد أن قيمته هي 0,971 وبالتلي تكون قيمة معامل التحديد 0,943 ، ونقوم الآن بتطبيق الصيغة العامة لحساب معامل التحديد كنسبة بين مجاميع المربعات لنرى أننا سوف نصل إلى نفس النتيجة،

أول خطوة نتبعها هي التعويض بالقيم المشاهدة للمتغير المستقل س في المعادلة السابقة لنحصل على القيم المقدرة صَ،

س	صَ = ٨,٢١١ + ٠,١٣٢٥س	صَ
١٨٥	٨,٢١١+ ٠,١٣٢٥× ١٨٥=	٣٢,٧١٥
١٩٠	٨,٢١١ + ٠,١٣٢٥× ١٩٠=	٣٣,٣٧٧
٢٣٠	٨,٢١١+ ٠,١٣٢٥× ٢٣٠=	٤٥,٠٠٠
٢٤٠	٨,٢١١ + ٠,١٣٢٥ × ٢٤٠=	٣٨,٦٧٦
٢٧٥	٨,٢١١+ ٠,١٣٢٥ × ٢٧٥=	٤٤,٦٣٦
٣٢٠	٨,٢١١ + ٠,١٣٢٥ × ٣٢٠=	٥٠,٥٩٦

ويظهر الجدول التالي التغيرات المختلفة ومربعاتها ومنه نلاحظ ما يلي:

1- مجاميع التغيرات الكلية، في العمود الثالث، والمفسرة بالعلاقـة بـين المتغيرين، في العمود الرابع، وغير المفسرة أو العشوائية، في العمود الخامس، كلها تساوي الصفر،

2- حاصل جمع مجموع مربعات التغيرات المفسرة ومجمـوع مربعـات التغيرات غير المفسرة يساوي مجموع مربعات التغيرات الكلية،

٢٥٨ = ١٤,٨٢٨ + ٢٤٣,١٧٢

3- معامل التحديد هو النسبة بين مجموع المربعات المفسر إلى مجمـوع المربعات الكلي

$$\text{معامل التحديد} = \frac{٢٤٣,١٧٢}{٢٥٨} = ٠,٩٤٣$$

وهي نفس القيمة السابقة،

4 – إن ارتفاع قيمة معامل التحديد بهذه الصورة وبالتالي انخفاض نسـبة التغيرات غير المفسرة (٥,٧%) تعني أن النموذج الذي يربط بـين سـعر الكتاب وعدد صفحاته يضمن إعطاء تنبؤات عالية الدقة لتقدير سعر الكتاب إذا علمنا عدد صفحاته،

ص	ص̂	ص − ص̄	ص̂ − ص̄	ص − ص̂	$(ص - \bar{ص})^2$	$(\hat{ص} - \bar{ص})^2$	$(ص - \hat{ص})^2$
٨٠	٣٢,٥٤٩	١٠−	٧,٤٥١−	٢,٥٤٩−	١٠٠	٥٥,٥١٧	٦,٤٩٨
٣٥	٣٣,٢٢٦	٥−	٦,٧٧٤−	١,٧٧٤	٢٥	٤٥,٨٨١	٣,١٤٦
٣٨	٣٨,٦٢٥	٢−	١,٣٥٥−	٠,٦٤٥−	٤	١,٨٣٥	٠,٤١٦
٤٢	٤٠	٢	٠	٢	٤	٠	٤
٤٥	٤٤,٧٤٢	٥	٤,٧٤٢	٠,٢٥٨	٢٥	٢٢,٤٨٢	٠,٠٦٦
٥٠	٥٠,٨٣٨	١٠	١٠,٨٣٨	٠,٨٣٨−	١٠٠	١١٧,٤٥٧	٠,٧٠٢
٢٤٠	٢٤٠	صفر	صفر	صفر	٢٥٨	٢٤٣,١٧٢	١٤,٨٢٨

٦-٧ معامل سبيرمان لارتباط الرتب

The Spearman Ranks Correlation Coefficient

ذكرنا أن معامل بيرسون للارتباط الخطي بين متغيرين يتطلب أن يكونا كميين. وإذا كان أحد المتغيرين أو كلاهما نوعي ترتيبي فإننا نقوم في هذه الحالة باستخدام معامل آخر للارتباط يطلق عليه معامل سبيرمان للارتباط أو معامل ارتباط الرتب. كذلك ينصح أيضا باستخدام معامل ارتباط الرتب في حالة البيانات الكمية التي تحتوي على بعض القيم الشاذة حيث يتأثر معامل بيرسون للارتباط بوجود مثل هذه القيم بينما يكون معامل ارتباط الرتب أقل تأثراً بها.

ولحساب معامل ارتباط الرتب نتبع ما يلي:

1- نعين رتبة لكل صفة أو لكل قيمة من قيم المتغير الأول وفقاً لترتيب تصاعدي (أو تنازلي) للصفات أو القيم.

2- نتبع نفس الأسلوب في تعيين رتب لصفات أو قيم المتغير الثاني.

3- نقوم بحساب الفروق المتناظرة بين رتب المتغيرين

ف = رتبة المتغير الأول – الرتبة المناظرة للمتغير الثاني.

4- نقوم بإيجاد مربع كل فرق من الفروق السابقة ونأخذ المجموع (مجـ ف^{٢}).

5- أخيراً نستخدم العلاقة التالية لحساب قيمة معامل سبيرمان لارتباط الرتب

$$ر = ١ - \frac{٦ مجـ ف^{٢}}{ن(ن^{٢}-١)}$$

يلاحظ أنه يمكن أيضاً الحصول على قيمة معامل سبيرمان للارتباط دون أخذ الفروق ومربعاتها وذلك عن طريق حساب معامل بيرسون للارتباط بين رتب المتغيرين بشرط عدم وجود تكرار لقيم أو صفات كل من المتغيرين.

مثال ٧-١٤

جمعت البيانات التالية لدراسة العلاقة بين عدد سنوات الخبرة وتقدير الكفاءة لعينة من عمال مصنع معين.

عدد سنوات الخبرة	تقدير كفاءة العامل
٨	ممتاز
٥	جيد جدا
٧	جيد
٣	مقبول
١	ضعيف

والمطلوب حساب الارتباط بين مستوى الخبرة والكفاءة.

الحل:

حيث أن متغير الكفاءة يكون وصفي ترتيبي وليس كمي، فإننا نستخدم معامل سبيرمان لقياس درجة الارتباط بين المتغيرين

سنوات الخبرة	تقدير كفاءة	رتب الخبرة	رتب الكفاءة	فروق الرتب	مربعات الفروق
٨	ممتاز	٥	٥	٠	٠
٥	جيد جدا	٣	٤	١-	١
٧	جيد	٤	٣	١	١
٣	ضعيف	٢	١	١	١
١	مقبول	١	٢	١-	١
مجـ				صفر	٤

يلاحظ في الجدول السابق:

1- عند تعيين رتب الخبرة أعطينا أقل القيم الرتبة الأولى ثم القيمة التالية لها في الصغر الرتبة الثانية وهكذا. وبالتالي لابد من اتباع نفس الأساس في تعيين رتب تقديرات الكفاءة حيث عينا الرتبة الأولى لأقل التقديرات والرتبة الثانية للتقدير التالي وهكذا.

2- إذا قمنا بتعيين الرتب بصورة صحيحة، فلا بد وأن نجد أن مجموع عمود الفروق يساوي صفر.

$$\text{معامل سبيرمان للارتباط} = ١ - \frac{٦ \times ٤}{٥(٢٥ - ١)}$$

$$= ١ - ٠,٢ = ٠,٨$$

ويعني هذا أنه يوجد اتجاه علاقة طردية قوية بين عدد سنوات الخبرة وتقدير الكفاءة.

وحيث أنه لا توجد قيم مكررة في عدد سنوات الخبرة ولا توجد صفة مكررة في تقديرات الكفاءة، فإننا سوف نحصل على نفس النتيجة السابقة إذا ما قمنا بحساب معامل بيرسون للارتباط بين رتب المتغيرين. فإذا استخدمنا س للدلالة على رتب الخبرة وص للدلالة على رتب الكفاءة، نحصل على المعلومات التالية

س	ص	س ص	$س^2$	$ص^2$
٥	٥	٢٥	٢٥	٢٥
٣	٤	١٢	٩	١٦
٤	٣	١٢	١٦	٩
٢	١	٢	٤	١
١	٢	٢	١	٤
١٥	١٥	٥٣	٥٥	٥٥

معامل بيرسمون للارتباط بين رتب المتغيرين

$$ر = \frac{٥٣ - \frac{١٥ \times ١٥}{٥}}{\sqrt{٥٥ - \frac{(١٥)^{٢}}{٥}}\ \sqrt{٥٥ - \frac{(١٥)^{٢}}{٥}}}$$

$$ر = \frac{٥٣ - ٤٥}{١٠} = ٠,٨$$

يلاحظ في معظم الأحيان عند حساب معامل ارتباط الرتب أن هناك بعض الصفات أو القيم يتكرر ظهورها أكثر من مرة وفي هذه الحالة نعين

لكل صفة أو قيمة مكررة متوسط الرتب التي تحتلها. فعلى سبيل المثال إذا أردنا تعيين رتب للقيم

٢	٤	٤	٥	٧	٧	٧	٨

فإن القيمة ٢ تأخذ الرتبة ١، القيمة ٤ تكرر ظهورها مرتين وتحتل الرتبتين الثانية والثالثة وبالتالي نقوم بإعطاء كل منهما الرتبة ٢,٥ . بعد ذلك تأتي القيمة ٥ لتحتل الرتبة الرابعة. وحيث أن القيمة ٧ تكرر ظهورها ثلاث مرات وتحتل الرتب ٥، ٦، ٧ فإننا نعين لكل منها متوسط الرتب الثلاث (٦). وأخيراً تحتل القيمة ٨ الرتبة الثامنة كما يتضح من الجدول التالي.

القيمة	٢	٤	٤	٥	٧	٧	٧	٨
الرتب	١	٢,٥	٢,٥	٤	٦	٦	٦	٨

مثال ٧-١٥

البيانات التالية تمثل تقديرات عينة من الطلبة في امتحانين للإحصاء والاقتصاد . والمطلوب حساب قيمة معامل الارتباط بين التقديرات في المادتين.

تقديرات الإحصاء	تقديرات الاقتصاد
جيد	جيد
جبد	مقبول
مقبول	جيد
مقبول	مقبول
ممتاز	جيد جدا
ضعيف	مقبول
ممتاز	ممتاز
جيد جدا	جيد

نقوم في هذا المثال بتعيين الرتب وفقا لترتيب تنازلي للتقديرات فنبدأ بإعطاء الرتب الأولى للتقدير ممتاز ثم الرتب التالية للتقدير جيد جدا وهكذا. يلاحظ أن التقدير ممتاز تكرر مرتين في تقديرات الإحصاء وبالتالي يحتل المرتبتين الأولى والثانية ويكون متوسط الرتب هو 1.5 . ويأخذ الطالب الذي حصل على جيد جدا الرتبة الثالثة وهكذا. وتظهر رتب المادتين كما في الجدول التالي

تقديرات الإحصاء	تقديرات الاقتصاد	رتب الإحصاء	رتب الاقتصاد	ف	$ف^2$
جيد	جيد	٤,٥	٤	٠,٥	٠,٢٥
جبد	مقبول	٤,٥	٧	-٢,٥	٦,٢٥
مقبول	جيد	٦,٥	٤	٢,٥	٦,٢٥
مقبول	مقبول	٦,٥	٧	-٠,٥	٠,٢٥
ممتاز	جيد جدا	١,٥	٢	-٠,٥	٠,٢٥
ضعيف	مقبول	٨	٧	١	١
ممتاز	ممتاز	١,٥	١	-٠,٥	٠,٢٥
جيد جدا	جيد	٣	٤	-١	١
مجـ				صفر	١٥,٥

معامل سبيرمان للارتباط

$$ر = ١ - \frac{٦ \times ١٥,٥}{٨(٦٤ - ١)}$$

$$= ١ - ٠,١٨٥ = ٠,٨١٥$$

من هذا يتضح وجود علاقة طردية قوية بين المتغيرين. أي أن الطالب الذي يحصل على تقدير مرتفع في الإحصاء من المتوقع أن يكون تقديره مرتفع أيضا في الاقتصاد والعكس صحيح.

مثال ٧-١٩

نقوم في هذا المثال بحساب قيمة معامل ارتباط الرتب من بيانات مثال ٧-٩ والذي سبق أن ذكرنا أن المشاهدة الأخيرة في بياناته تعتبر قيمة غير

طبيعية وكان لاستبعاد هذه القيمة أثر كبير حيث ارتفعت قيمة معامل بيرسون للارتباط من ٠,٧٢ إلى ٠,٩٧ . ونعيد كتابة البيانات في الجدول التالي ثم نقوم بحساب قيمة معامل ارتباط الرتب دون استبعاد المشاهدة الأخيرة.

س	٣	٤	٤	٤	٥	٦	٨
ص	٥	٦	٧	٦	٨	١٠	٨

الحل

س	ص	رتب س	رتب ص	ف	ف^2
٣	٥	١	١	٠	٠
٤	٦	٣	٢,٥	٠,٥	٠,٢٥
٤	٧	٣	٤	١-	١
٤	٦	٣	٢,٥	٠,٥	٠,٢٥
٥	٨	٥	٥,٥	٠,٥-	٠,٢٥
٦	١٠	٦	٧	١-	١
٨	٨	٧	٥,٥	١,٥	٢,٢٥
مجـ				صفر	٥

$$ر = ١ - \frac{٥ \times ٦}{٧(٤٩ - ١)}$$

$$= ١ - ٠,٠٨٩ = ٠,٩١$$

يتضح من هذا المثال أن قيمة معامل سبيرمان لارتباط الرتب لم تتأثر بدرجة كبيرة بوجود القيمة المتطرفة مثل معامل بيرسون للارتباط.

تمارين الفصل السابع

١-) جمعت البيانات التالية لقياس العلاقة بين حجم الإنفاق الشهري علـــى السلع الغذائية والدخل الشهري لعينة من الأسر

الإنفاق	540	330	680	420	720	480	390	440
الدخل	800	560	900	640	1200	680	620	760

احسب معامل بيرسون للارتباط وعلق النتيجة.

٢-) أعد حل التمرين السابق بعد طرح 330 كوسط فرضي من قيم الإنفاق و 560 كوسط فرضي من قيم الدخل ثم قسمة جميع النواتج على 10.

٣-) فيما يلي بيانات بتقديرات عينة من طلبة السنة الأولى في كلية التجارة والنسب المئوية لدرجاتهم في امتحان الثانوية العامة

التقدير	جيد	مقبول	جيد	جيـــد جدا	ضعيف	مقبول	جيد
النسبة	%87	%82	%84	%90	%80	%78	%84

احسب مقياس ملائم للارتباط لقياس قوة العلاقة بين المتغيرين وعلق على النتيجة.

الجـداول

$\Phi(z) = P(Z \leq z)$ جدول التوزيع الطبيعي المعياري

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Selected Normal Percentiles

P	.90	.95	.975	.99	.995
Z	1.2816	1.6449	1.96	2.3263	2.5758

جدول توزيع *

Degrees Of Freedom	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$T_{0.005}$
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
∞	1.282	1.645	1.960	2.326	2.576

جدول توزيع χ^2

Degrees Of Freedom	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	2.7055	3.8415	5.0239	6.6349	7.8794
2	4.6052	5.9915	7.3778	9.2104	10.5965
3	6.2514	7.8147	9.3484	11.3449	12.8381
4	7.7794	9.4877	11.1433	13.2767	14.8602
5	9.2363	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5475
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5345	20.0902	21.9549
9	14.6837	16.9190	19.0228	21.6660	23.5893
10	15.9872	18.3070	20.4832	23.2093	25.1881
11	17.2750	19.6752	21.9200	24.7250	26.7569
12	18.5493	21.0261	23.3367	26.2170	28.2997
13	19.8119	22.3620	24.7356	27.6882	29.8193
14	21.0641	23.6848	26.1189	29.1412	31.3194
15	22.3071	24.9958	27.4884	30.5780	32.8015
16	23.5418	26.2962	28.8453	31.9999	34.2671
17	24.7690	27.5871	30.1910	33.4087	35.7184
18	25.9894	28.8693	31.5264	34.8052	37.1564
19	27.2036	30.1435	32.8523	36.1908	38.5821
20	28.4120	31.4104	34.1696	37.5663	39.9969
21	29.6151	32.6706	35.4789	38.9322	41.4009
22	30.8133	33.9245	36.7807	40.2894	42.7957
23	32.0069	35.1725	38.0756	41.6383	44.1814
24	33.1962	36.4150	39.3641	42.9798	45.5584
25	34.3816	37.6525	40.6465	44.3140	46.9280
26	35.5632	38.8851	41.9231	45.6416	48.2898
27	36.7412	40.1133	43.1945	46.9628	49.6450
28	37.9159	41.3372	44.4608	48.2782	50.9936
29	39.0875	42.5569	45.7223	49.5878	52.3355
30	40.2560	43.7730	46.9792	50.8922	53.6719
40	51.8050	55.7585	59.3417	63.6908	66.7660

$\alpha = 0.05$

v_1 / v_2	9	10	11	12	15	20	24	30
	Numerator Degrees of Freedom							
1	240.54	241.88	242.98	243.90	245.95	248.02	249.05	250.10
2	19.38	19.40	19.40	19.41	19.43	19.45	19.45	19.46
3	8.81	8.79	8.76	8.74	8.70	8.66	8.64	8.62
4	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75
5	4.77	4.74	4.70	4.68	4.62	4.56	4.53	4.50
6	4.10	4.06	4.03	4.00	3.94	3.87	3.84	3.81
7	3.68	3.64	3.60	3.57	3.51	3.44	3.41	3.38
8	3.39	3.35	3.31	3.28	3.22	3.15	3.12	3.08
9	3.18	3.14	3.10	3.07	3.01	2.94	2.90	2.86
10	3.02	2.98	2.94	2.91	2.85	2.77	2.74	2.70
11	2.90	2.85	2.82	2.79	2.72	2.65	2.61	2.57
12	2.80	2.75	2.72	2.69	2.62	2.54	2.51	2.47
13	2.71	2.67	2.63	2.60	2.53	2.46	2.42	2.38
14	2.65	2.60	2.57	2.53	2.46	2.39	2.35	2.31
15	2.59	2.54	2.51	2.48	2.40	2.33	2.29	2.25
16	2.54	2.49	2.46	2.42	2.35	2.28	2.24	2.19
17	2.49	2.45	2.41	2.38	2.31	2.23	2.19	2.15
18	2.46	2.41	2.37	2.34	2.27	2.19	2.15	2.11
19	2.42	2.38	2.34	2.31	2.23	2.16	2.11	2.07
20	2.39	2.35	2.31	2.28	2.20	2.12	2.08	2.04
21	2.37	2.32	2.23	2.25	2.18	2.10	2.05	2.01
22	2.34	2.30	2.26	2.23	2.15	2.07	2.03	1.98
23	2.32	2.27	2.24	2.20	2.13	2.05	2.01	1.96
24	2.30	2.25	2.22	2.18	2.11	2.03	1.98	1.94
25	2.28	2.24	2.20	2.16	2.09	2.01	1.96	1.92
26	2.27	2.22	2.18	2.15	2.07	1.99	1.95	1.90
27	2.25	2.20	2.17	2.13	2.06	1.97	1.93	1.88
28	2.24	2.19	2.15	2.12	2.04	1.96	1.91	1.87
29	2.22	2.18	2.14	2.10	2.03	1.94	1.90	1.85
30	2.21	2.16	2.13	2.09	2.01	1.93	1.89	1.84

جدول توزيع F

$\alpha = 0.05$

v_2 \ v_1	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27

$\alpha = 0.01$

v_2 \ v_1	9	10	11	12	15	20	24	30
	Numerator Degrees of Freedom							
1	6022.40	6055.93	6083.40	6106.68	6156.97	6208.66	6234.27	6260.35
2	99.39	99.40	99.41	99.42	99.43	99.45	99.46	99.47
3	27.34	27.23	27.13	27.05	26.87	26.69	26.60	26.50
4	14.66	14.55	14.45	14.37	14.20	14.02	13.93	13.84
5	10.16	10.05	9.96	9.89	9.72	9.55	9.47	9.38
6	7.98	7.87	7.79	7.72	7.56	7.40	7.31	7.23
7	6.72	6.62	6.54	6.47	6.31	6.16	6.07	5.99
8	5.91	5.81	5.73	5.67	5.52	5.36	5.28	5.20
9	5.35	5.26	5.18	5.11	4.96	4.81	4.73	4.65
10	4.94	4.85	4.77	4.71	4.56	4.41	4.33	4.25
11	4.63	4.54	4.46	4.40	4.25	4.10	4.02	3.94
12	4.39	4.30	4.22	4.16	4.01	3.86	3.78	3.70
13	4.19	4.10	4.02	3.96	3.82	3.66	3.59	3.51
14	4.03	3.94	3.86	3.80	3.66	3.51	3.43	3.35
15	3.89	3.80	3.73	3.67	3.52	3.37	3.29	3.21
16	3.78	3.69	3.62	3.55	3.41	3.26	3.18	3.10
17	3.68	3.59	3.52	3.46	3.31	3.16	3.08	3.00
18	3.60	3.51	3.43	3.37	3.23	3.08	3.00	2.92
19	3.52	3.43	3.36	3.30	3.15	3.00	2.92	2.84
20	3.46	3.37	3.29	3.23	3.09	2.94	2.86	2.78
21	3.40	3.31	3.2[illegible]	3.17	3.03	2.88	2.80	2.72
22	3.35	3.26	3.18	3.12	2.98	2.83	2.75	2.67
23	3.30	3.21	3.14	3.07	2.93	2.78	2.70	2.62
24	3.26	3.17	3.09	3.03	2.89	2.74	2.66	2.58
25	3.22	3.13	3.06	2.99	2.85	2.70	2.62	2.54
26	3.18	3.09	3.02	2.96	2.81	2.66	2.58	2.50
27	3.15	3.06	2.99	2.93	2.78	2.63	2.55	2.47
28	3.12	3.03	2.96	2.90	2.75	2.60	2.52	2.44
29	3.09	3.00	2.93	2.87	2.73	2.57	2.49	2.41
30	3.07	2.98	2.91	2.84	2.70	2.55	2.47	2.39

جدول توزیع F

$\alpha = 0.01$

v_1 / v_2	1	2	3	4	5	6	7	8
	Numerator Degrees of Freedom							
1	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95
2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17

المحتويات







الناشر
مكتبة الوفاء القانونية

٣٧٣٨٨٢٢ ٠١٠ ٠٠٢

تليفاكس
٠٠٢٠٣/٥٤٠٤٤٨٠
الإسكندرية